The Scientific Basis of Medicine Annual Reviews 1966

THE ATHLONE PRESS
1966

CONTENTS

I. The Ethical Basis of Medical Science 1
Sir Robert Platt, BT., M.SC., M.D., F.R.C.P.
Department of Medicine, Manchester Royal Infirmary

II. Environmental Radioactivity and its Biological Significance 16
W. V. Mayneord, C.B.E., D.SC., F.INST.P., F.R.S.
Emeritus Professor of Physics as applied to Medicine, The University of London

III. Ultrasonics in Clinical Diagnosis 38
P. N. T. Wells, M.SC.
Department of Medical Physics, Bristol Royal Hospital

IV. Sophisticated Laboratory Animals 54
W. Lane-Petter, M.B., B.CHIR.
M.R.C. Laboratory Animals Centre, Carshalton

V. Demography and the Cancers 71
R. A. M. Case, M.D., PH.D.
Department of Environmental Pathology, Institute of Cancer Research, Royal Cancer Hospital, London

VI. The Age of Reaching Puberty: Some Associated Factors and Some Educational Implications 91
J. W. B. Douglas, B.SC., B.M., B.CH.
M.R.C. Unit for the Study of Environmental Factors in Mental and Physical Illness, London School of Economics and Political Science

VII. Disturbances of the Immunoglobulins 106
J. R. Hobbs, B.SC., M.D., M.R.C.P., M.C. PATH.
Department of Pathology, Postgraduate Medical School of London

VIII. Rheumatoid Factor 128
JOHN R. SQUIRE, M.D., F.R.C.P.
Department of Experimental Pathology, University of Birmingham

IX. Histochemical Techniques in the Study of the Cell Biology of Dental Tissues 146
A. R. TEN CATE, B.SC., PH.D., B.D.S.
Department of Anatomy, Guy's Hospital Medical School, London

X. Staphylococci Endemic in Hospitals 157
M. T. PARKER, M.D.
Cross Infection Reference Laboratory, Central Public Health Laboratory, London

XI. Antiviral Chemotherapy 174
D. J. BAUER, M.B., B.CHIR., PH.D., M.R.C.P.
Wellcome Laboratories of Tropical Medicine, London

XII. Dilute and Concentrated Urine 192
J. N. MILLS, D.M.
Department of Physiology, University of Manchester

XIII. Long-term Dialysis as a Substitute for Kidney Function 201
STANLEY SHALDON, M.D., M.R.C.P.
Renal Unit, Royal Free Hospital, London

XIV. The Transport of Sodium and Potassium across Cell Membranes 217
I. M. GLYNN, M.B., B.CHIR., PH.D.
Physiological Laboratory, University of Cambridge

XV. Formation and Drainage of the Cerebrospinal Fluid 238
H. DAVSON, D.SC.
Department of Physiology, University College London

XVI. Nervous Control of Limb Blood Flow 260
I. C. RODDIE, D.SC., M.D., M.R.C.P.I.
Department of Physiology, The Queen's University, Belfast

XVII. Distribution of Pulmonary Blood Flow 279
JOHN B. WEST, M.D., PH.D.
Department of Medicine, Postgraduate Medical School of London

XVIII. Tissue Oxygen Requirements 300
I. S. LONGMUIR, M.B., B.CHIR.
Department of Biochemistry, Institute of Diseases of the Chest, London

XIX Thyroid and Central Nervous Development 317
J. T. EAYRS, PH.D., D.SC.
Department of Physiology, Institute of Psychiatry, London

'The Scientific Basis of Medicine': Index of authors and papers 341

I

The Ethical Basis of Medical Science

SIR ROBERT PLATT

THE progress of medical science has greatly increased the doctor's responsibility, for he has so much more control over the outcome of disease. Life and death often depend upon his skill in a way which they rarely did 30 years ago. Two examples will suffice: in serious infective diseases such as meningitis or bacterial endocarditis success in treatment depends upon prompt recognition, thus early and accurate diagnosis takes on a new importance. Surgery on the open heart requires clinical judgement in the selection of cases as well as efficiency of the highest order in the surgical team. A mistake may prove fatal at any stage from initial assessment to post-operative care. But the day-to-day practice of medicine is only rarely a matter of life and death and in any case a large number of deaths nowadays are either the immediate result of the internal combustion engine or the inevitable outcome of old age or of chronic diseases for which science has still no cure and I would like therefore to refer to the type of decision which medical science has posed upon the family doctor in the management of common illness. Young men who have never ventured beyond the nursery of the academic medical unit are apt to see a supposed conflict between the accurate and logical behaviour of the medical scientist on the one hand and the slipshod methods of an over-worked general practitioner on the other who finds it easier to write a prescription than to think. But in fact many of the decisions made in general practice are at least as rational as those made in hospital, and sometimes rest on firmer ground.

The supposed conflict derives from the need of the practitioner to act quickly in the type of case which the specialist rarely sees. I will try to illustrate this by a duologue between two responsible doctors one of whom, a general practitioner, has just been called to a child with acute earache. He says: 'I give penicillin to all these cases.' His specialist friend says: 'But in your practice you have direct access to a good laboratory; you mean that you have no bacteriological control of what you are doing?' 'Bacteriology would be of very little help. I can't take a swab from the middle ear until the drum has perforated and that is precisely what I want to prevent. Even if I take throat or nasal swabs the drum will probably perforate, if it is going to do so, before I get the report.' 'But even if the drum perforates it will quickly heal again, especially if the right antibiotic is given at that stage.' 'Yes, but not without some loss of hearing in some cases, which will probably be permanent; and I also think I shorten the period of pain and prevent complications such as mastoiditis, cerebral abscess and meningitis, which we used to see but which have now almost completely disappeared. You may not realize that only 1 per cent of cases of otitis media now ever reach hospital at all.' 'You don't think you should be selective and try to judge the kind of case in which complications are likely to occur?; or try to distinguish between virus and bacterial infections?' 'Unfortunately there is no very helpful evidence on which such judgements can be made; therefore I must treat all cases, provided, of course, that the antibiotic I am using is not a dangerous one, which would be unjustified in an illness which usually recovers.' 'But there is still the danger of producing penicillin-resistant organisms and penicillin sensitivity.' 'This I admit, but the fact is that penicillin-resistant organisms have developed much more in hospital than amongst patients in general practice. Penicillin sensitivity is rare, especially if oral penicillin is used.' 'If you really think that penicillin should be given efficiently to all cases as early as possible, why not give it by injection, why trouble about a pinprick where efficiency is concerned?' 'Because I might not only build up penicillin-intolerance but doctor-intolerance which might make it almost impossible to treat this child for any future illness. May I now ask you a question? When a patient is admitted to hospital with suspected bronchopneumonia, do you give antibiotics indiscriminately without waiting for laboratory reports?' 'Yes.'

The argument, of course, could have been prolonged, and a similar case might be made out for the indiscriminate use of antibiotics in all cases of tonsillitis. It is difficult to prove that any harm has come from this, whereas the complications of tonsillitis, which include acute nephritis and rheumatic fever, have become much rarer since penicillin was put into the practitioner's hands. Moreover the widespread use of antibiotics has altered the pattern of infective disease and although there is still the possibility that it may alter for the worse (indeed there is some recent evidence of this with regard to dysenteric diseases), up to the present most of the changes have been beneficial. Pneumococcal pneumonia is now a rarity and pulmonary tuberculosis shows signs of dying out altogether. In hard cash it is easy to assess the cost of antibiotics but impossible to estimate how much they save.

Computer Diagnosis

Computers will not take personal responsibility from our shoulders any more than biochemistry or radiology have done, though like other tools of medical science they may give new and refined data on which judgements can be made. Medical research is finding increasing use for computers, but a computer is essentially a highly efficient statistical machine and suffers from the two inherent defects of statistics: that they can never be more reliable than the data on which they are based, and that they can help very little with the individual case. In addition to their value for medical research, computers may find a place in the evaluation of symptoms and signs, but those who suggest that they will eventually take over the diagnostic problems of the doctor have not, I think, adequately considered what the problems of diagnosis really are.

To feed into a computer a list of the signs and symptoms and expect the answer is a naïve modern version of the hypochondriac of former years who made his own diagnosis (erroneously of course) by looking up his symptoms in the medical dictionary. The essential difficulties of diagnosis are usually in the ascertainment of the data and the evaluation of their relative importance and accuracy. All symptoms are subjective, and can only be elucidated by careful collaboration between doctor and patient. What the patient will tell you depends on whether you seem to him or her to be sympathetic and interested, or cold and forbidding, or worse still in a hurry to get rid of him (or her). I challenge any computer to elucidate the problems of some of my patients and their relevance to the symptoms. Data from some of the most interesting ones (speaking shall we say psycho-socially) fed into a computer by three different physicians would give three different answers.

OVER-INVESTIGATION

The syndrome of over-investigation is one of the byproducts of science and merits inclusion amongst the doctor-induced, or iatrogenic diseases of our time, though it has not yet been studied by departments of preventive medicine. No better example could be given of a situation in which this may develop than our present dilemma over the relation of renal artery stenosis and unilateral pyelonephritis to hypertensive disease. A hundred patients, after an anxious period of delay, are subjected to the expensive, time-consuming and rather frightening technique of aortography and twenty of these to the discomforts and indignities of ureteric catheterization and perhaps ten of them to surgery, and out of the hundred perhaps three at the most are a real success, cured of their disease. Of the 97 remaining, the surgical failures are dejected and may by now have lost their employment. In the United States of America they would also have lost their savings. Those not operated on feel discarded as incurable or are left in doubt about their future, and become confirmed in invalidism unless they are unusual personalities handled by exceptional doctors. Yet it seems that the investigations must continue, not only for the occasional success, but in the hope that better prediction of the end-result will soon become a possibility. In the meantime it is my belief that no doctor need feel guilty for not subjecting all his hypertensive patients to the devices of medical science. Because the successes are described not only in scientific journals but probably also on television, there is a danger that anyone treating high blood pressure without full investigation is looked upon as primitive, and so what should be a procedure reserved for selected cases in special centres becomes a matter in which all hospitals try to compete, and because the harm done to the failures is psychological it is not measurable and therefore not usually recorded.

ORGAN TRANSPLANTION

A similar dilemma occurs in much more acute form at the present time where transplantation of a kidney is contemplated. Recent figures showed that of 148 cases of renal grafting between individuals who were not identical twins, only three transplants survived for two years (Woodruff, 1964), and so in addition to subjecting a seriously ill patient to a dangerous operation with only a minute chance of success, we have the difficult ethical situation of removing an important organ from a healthy person or of waiting hopefully to see a suitable donor die so that his kidney can be snatched almost at the moment of death. At the present time it would seem that the operation is only justifiable if it is done with a view to the future as well as the present; that is in a unit which is capable of carrying out original work, the results of which may throw light on the all-important questions of the survival of renal grafts. But to come to this conclusion is in fact a new departure in the field of medical ethics for it brings the question of eventual benefit to others into a decision about the treatment and selection of individuals, which could be a dangerous precedent. There is no lack of willing collaborators amongst patients and their relatives and the temporary successes are so dramatic and well publicized that in renal grafting even more than in hypertension the danger arises that a procedure still at the stage of research and development attracts a prestige value out of keeping with its practical use at the present time.

SELECTION OF PATIENTS

The problems I have referred to involve the selection of patients, in the sense of finding those in which the procedure has some chance of success, but a new problem of selection is now emerging almost for the first time—the problem of who shall benefit from a procedure of known efficacy in which only a limited number of patients can be catered for. The obvious example is recurrent dialysis in chronic renal failure, for every patient thus kept alive adds to the numbers which may require twice-weekly dialysis for perhaps thirty or more years. Perhaps automation in industry will leave enough people unemployed to train as technicians in some vast and expanding dialysis centres, but the cost will be considerable. Such dilemmas of selection have occasionally occurred before, but only temporarily, when some new and life-saving drug has been in short supply. Bernard Shaw was well before his time in the *Doctor's Dilemma.*

CONTROLLED TRIALS IN HUMAN EXPERIMENT

On the subject of human experiment much has been written in recent years (some of it by myself) and I only intend to make some points in which my emphasis may be a little different from that of others. The Medical Research Council (1964) has recently set out in revised form a code of behaviour which will help to guide investigators of the present and the future, but like most codes of behaviour on this subject it starts with a tacit assumption that the experiment contemplated is a good one from a scientific point of view, that it is likely to give the answer to the question, and that the question is one which is worth asking. In that case the only consideration is whether and in what circumstances the experiment is justified. But not all the experiments carried out in modern medical departments are good ones, not because they are dangerous, nor because they are statistically unsound—in a purely numerical sense (Bradford Hill and his colleagues have largely weaned us from those early fallacies), but because we underestimate the magnitude of biological variation, neglect the emotional impact of the experiment on the patient, seek to measure the immeasurable and design a controlled trial in cases where controls are unnecessary, impossible to match or otherwise inappropriate. A controlled trial is at best a method of assessment. It does not give rise to new discovery or hypothesis.

I was recently involved in a television interview in which I was asked what I considered to be the most beneficial medical achievement of my time. With very little hesitation and no prior thought, I answered: 'The conquest of pulmonary tuberculosis', for this was a disease which meant death to young people and which ran in families and which didn't kill you very quickly or without suffering, a disease which had accounted for the early demise of many persons of great talent, such as the Brontë family and Chopin and Keats, a disease so full of tragedy that it could be the theme of Puccini's luscious triumph of sentimentality, *La Bohème*. Only twenty years ago the X-ray evidence of tuberculosis in a young person had dreadful significance.

The cure of pulmonary tuberculosis was not discovered by clinical physiologists measuring lung function and expressing it in numerical terms, but in the laboratory by trial and error, by observations which could only have been planned on the prior knowledge of penicillin. Controlled trials were later made to assess the full efficacy and limitations of streptomycin, but these came after its discovery and when its anti-tuberculous action was already clear. Neither was the discovery of penicillin or insulin or the cure of pernicious anaemia based on the measurement of parameters or on the controlled trial. Nor did the benefits of mitral valvotomy need demonstration by the random selection of human guinea-pigs. Of course I am quoting some of the great successes for which statistical proof would be an act of supererogation. It is in the more marginal situations such as the value (if any) of long-term anticoagulants in coronary heart disease that the controlled trial is necessary to determine that one or other treatment will save a higher proportion of the lives involved, but it is often in just these cases that the number of variables defeats the experiment. The recent large-scale trial of long-term anticoagulant treatment in this country has been a dubious success for this reason (Reports to Med. Res. Council, 1959, 1964). It is unlikely that the trial of steroids in nephrosis will be more fruitful.

Bradford Hill (1963) has rightly made the point that a controlled trial is never justifiable unless the doctor is genuinely unconvinced as to whether the treatment is effective, or better than some other treatment. A comparison between medical and surgical treatment of serious disease by random selection of cases must usually be discarded on ethical grounds. A controlled trial of sympathectomy in hypertension for instance could never have passed the planning stage because no clinician would have relinquished his responsibility for judging an individual patient's mental or physical suitability to withstand a serious operation.

I doubt if the use of the placebo in a controlled trial is ever justifiable without the full consent and co-operation of the patient, which practically limits placebo trials to investigations on the treatment of symptoms rather than diseases unless the disease is either trivial or very chronic, but this difficulty can often be overcome by comparing one treatment with another. In the trial of long-term anticoagulant treatment to which I have referred the patients were randomly divided into two groups, those which had efficient therapy and those which had anticoagulants in such small doses as to be deliberately inefficient. This I once described as a double-blind trial of medical ethics, the doctors deceiving themselves that they were not deceiving the patients. A doctor-patient relationship based on deception is bad for both parties.

SUBJECT FOR RESEARCH

Some important discoveries and a great deal of knowledge of physiological function in health and disease have been acquired by making accurate measurements under varied environmental conditions. The classical experiments of McCance (1936) on salt deficiency were an outstanding example of this kind of discovery, and measurements of cardiac output, pulmonary artery pressure and respiratory function have contributed importantly to other discoveries and to the understanding of the processes of disease, but in the design of some experiments there is a danger of creating a situation so complex and artificial that the measurements cannot be interpreted in terms of any natural situation, especially if the parameter being measured, like blood pressure or diuresis, is liable to be affected by the emotional state of the patient. I also wonder whether the passion for measurement and for the design of an experiment, invested as it is with the respectability of being scientific has not impeded the progress of that kind of research which has to be based on observation and hypothesis, not only in fields such as psychology and psychiatry where hard facts are difficult to come by, but also in the study of the fascinating symptoms and signs of bodily illness, John Hunter said: 'Why think, why not do the experiment?' I am not the first to ask of some researchers of the present day: 'Why experiment? Why not think?'[1]

It must be admitted that there is a large area in which there is as yet no easy or quick way of testing hypotheses and most of us shrug our shoulders and say the field is not ripe for development, instead of seeing it as a challenge. But even without proof, a satisfactory hypothesis if it explains the observed facts will gradually gain acceptance. We do not need to go outside the field of human biology and organic disease to find plenty of unanswered questions. We know almost nothing about hunger, for instance, or the lack of it which is anorexia, though we know a little about thirst. We do not know why some people's hands are hot and others are cold, or why women who seem equally efficient biologically show such variation in what are colloquially called their vital statistics. If we want to know how ill a patient is we still go to the bedside and not to the laboratory. The moist palms and the anxious face remain the best indication that the duodenal ulcer is still bleeding, but we don't really know why.

The study of some of these and similar phenomena might be more fruitful and a lot more interesting than devising new electronic devices for measuring unimportant phenomena. Nevertheless much important research depends on the timely exploitation of new techniques.

When we look at the great discoveries of our time we realize that some of them seem to have depended on chance observations and a man who has made a significant observation fortuitously and wants to follow up its potentialities must usually be worth supporting. His trouble often is that he either will not or cannot find the kind of collaborator he needs to make good use of his observation. Its further elucidation may depend on knowledge and technique which he does not possess. It is not respectable to believe that great discoveries have resulted entirely from good luck, so we speak of the prepared mind, which is able to discern the importance of a chance discovery which might have passed unnoticed by lesser men, but there have been examples in which the prepared mind was prepared to wait a long time before making any use of the discovery, and we must not shut our eyes to the importance of the unprepared mind, or perhaps we should say the mind which has not been conditioned to a conventional channel of thought, the man who brings mathematics into clinical medicine (or into botany as Mendel did), or the techniques of engineering and lubrication into orthopaedic surgery.

---

[1] Sir Theodore Fox used this antithesis, but neither he nor I can find the reference.

Research and Medical Education

Thoughts on trends in medical research lead to the reflection that the study of medicine since it ceased to be an education for learned men followed by an apprenticeship, and became a rather insecure mixture of art and applied science has always been uneasy in the academic atmosphere. I wonder whether we should not try the experiment which has been suggested, of siting one of our new medical schools alongside a College of Advanced Technology, in an atmosphere in which arts, skills, crafts and techniques are respectable and science is the servant, not the master, and education is acknowledged, without apology, to be vocational. After all we *are* trying to train doctors. Physicians, surgeons, pharmacologists and physiologists would have as their colleagues some of the leaders in the application of scientific techniques to industry.

I can imagine the horror with which such a suggestion will be received in some quarters. It will be criticized as an attempt to produce second-rate doctors and lower the status of a great and learned profession by all those both in and out of medicine whose traditional and acquired habits of thought lead them to look down upon any one who actually knows how to do something. After all, in the good old days gentlemen didn't really need to.

This attitude of mind is very dangerous to the future of our country if we are to maintain its reputation in the conception and exploitation of new ideas. The fact that so many important advances in medicine have come from the drug firms shows that medical research is by no means dependent on the academic atmosphere of a university department. Such a school of medicine as I suggest would have a hard fight to get the right men and the right status and the right students, indeed it is almost certain that it would be prevented by prejudice from doing so, but what a triumph it would be if it were an unqualified success, collaborating with industry in the production of new drugs and the perfecting of prosthetic operations, and perhaps even diverting into medicine some of the psychologists whose talents are at present squandered in the meretricious studies of the advertising industry, whose outstanding and brilliant researches into what makes people buy which, could most rewardingly be applied to why patients complain of what.

Modern medicine and surgery has come to rely increasingly on all kinds of technologies. Surely the kind of staff which a college of technology attracts could be of great value to its further development. It would be sad indeed if this had the effect of dehumanizing medicine still further but the very fact that learning was directed towards creating good doctors should prevent this. The academic departments in clinical subjects which have mostly been set up since the Goodenough report have had many good effects but it cannot be said that their influence in bringing humanism to medicine has been conspicuous. Some would say quite the reverse. As soon as you intellectualize your subject you begin to escape from some of its practical implications. The only important recent attempt to study communications and morale in hospitals, and their effect on patients, has come from a department of the Manchester College of Science and Technology.

Commercial Ethics and Medicine

Many great advances in therapy have, as I have said, come from the big drug firms and the expensive stage of development of a discovery of possible value in therapeutics is largely a matter which has been taken over by the pharmaceutical industry. Work requiring hundreds of chemists, technicians, engineers, rabbits and guinea-pigs, is needed at this stage such as no university at present has the means to support. From the ethical point of view the integrity of the big drug firms has been exemplary within the context of a commercial society and if they are spending millions of pounds on research they must be allowed their profits and their advertising methods, especially when one considers the great financial risks which they run. Assuming that at the present time the commercial society is likely to continue, then within the accepted range of sales promotion, advertising has on the whole been ethical and often positively helpful to those who are looking for new and improved ways of treating symptoms or diseases. Note that I distinguish between the great drug firms who are actively engaged in research, and those who are only concerned in marketing new mixtures of old drugs, and 'elegant' preparations of doubtful value.

There have of course been some cases in which the warning of serious side-effects seems to have been inadequate. Drugs which may, however rarely, cause fatal side-effects, such as aplastic anaemia, should never be used or recommended in cases of trivial illness; but on the whole, warnings have been promptly heeded and accepted. In November 1961 Dr. Lenz of Hamburg drew attention to the possible association of thalidomide with the birth of limbless babies. The drug was withdrawn from the British market on December 2nd of the same year. Only a few days ago I received a notice that manufacturers were withdrawing a drug 'pending further investigation' because it may have been responsible for ten cases of jaundice. Ten cases of jaundice seems rather trivial compared with about 25,000 deaths annually in Britain from cancer of the lung. Yet the evidence against the cigarette is overwhelming, is accepted by scientists all over the world and by the three major independent bodies which have examined it, namely the Medical Research Council, the Royal College of Physicians and the committee set up by the Surgeon General of the United States. But representatives of the tobacco manufacturers appear on television screens and publicly say they are unconvinced, 'more research is required'. This, though it may be genuine on their part, is surely an open invitation to smokers to reject the evidence, even if (and this is difficult to believe) the manufacturers did not intend it to be so. No one expects them to withdraw cigarettes from the market. There would be an outcry from a large section of the public if they did, but how much more dignified and acceptable it would have been and how much more helpful for those concerned for the public health if the manufacturers had joined together and announced their reasons for discontinuing the advertisement of cigarettes 'pending further investigation'. Third in importance only to the control of pulmonary tuberculosis and malaria, I would place the revelation that cancer of the lung is (nearly always) a preventible disease amongst the great discoveries of scientific medicine.

Medical Research and the Community

One of the problems of the ethics of human experiment is that we are not now merely concerned with the possible dangers of new treatment for the individual but of a new preventive for the community. We have to ask the parents of children to take risks not only because of the (now) remote chance of their child developing poliomyelitis or tuberculosis, but also for the benefit of the public health and the control of infection. This of course has been a problem ever since vaccination for smallpox became a practical possibility, but has become more acute with the accelerated progress of discovery. Speaking as a grandfather, one of the most difficult decisions I ever had to make was whether to agree to the trial of poliomyelitis vaccine in this country, I being a member of the Medical Research Council at that time and there having been recent accidents in the United States.

Medical progress has already had other and far-reaching effects on the community. The control of malaria and other widespread diseases and the reduction in infant mortality have much increased the rate of growth of the population in communities which have not yet solved their basic problems of nutrition. This is a subject on which much has been written and on which I have nothing new to say.

Modern discoveries in contraception may eventually be the answer to the threat of over-population but it seems inevitable that contraceptive methods will be used by the intelligent long before they are effectively used by the mass of the unskilled. In any case it will take a century or more to know the full effects of efficient and acceptable contraceptive techniques on the human race. To me it seems wrong, though in some cases temporarily expedient, that a woman's physiological mechanisms should be continuously disturbed by taking a contraceptive pill for three weeks in every month in order to guard against an unwelcome pregnancy. The ideal contraceptive pill would be taken either by the male immediately before intercourse or by the female immediately afterwards. Of course sexual behaviour would be altered by such a discovery but it is already being altered by the present contraceptive pill. In a small town in Tasmania which I recently visited, I was told that the maternity hospital now has empty beds, and that the High School girls call at the chemist's shop for their contraceptive pills at lunch time when the (male) senior pharmacist is off duty. I am sure that this small Tasmanian town is not unique in this respect.

The contraceptive pill is a product of medical research, and its social and emotional effects deserve consideration by medical scientists as well as its purely physiological effects. Man's code of sexual behaviour, which of course is not always adhered to, has evolved because of the survival value of a monogamous social structure in a species in which the young take many years to reach maturity. The code of sexual behaviour might have been differently written if it had evolved in a society to whom the perfect contraceptive was known. It is already clear that the sexual experience of adolescents of both sexes is less restricted than it was in former generations. I see no likelihood that this trend will be reversed. Whether it is for good or ill will depend on how it affects the stability of the emotional relationships which develop between young people, in other words whether it makes for better marriages or worse. Personally I think that the creative force which impels young men and women to found a home and bring up a family will prevail. The mating and the nesting instincts are closely interwoven. The availability of housing may be more important for the stability of marriage than the question of premarital intercourse, and the racket in rents and land values may at length be seen by a more enlightened generation as a much greater sin against society.

The ideal contraceptive will be a boon to the married and reduce the alarming rate of births to the unmarried. Young people want guidance and discussion, not dogma; they need positive incentives to a new morality based on the importance of home, marriage and family. This they should learn from parents, teachers and doctors. When I was a medical student no one ever spoke to us about these things. Forty years of progress have made little difference to this unfortunate reticence. Yet if preventive medicine and public health are the concern of doctors, then so is the mental health of society. A training for medicine which is obsessed with physical science and almost ignores human emotion and behaviour has lost its sense of proportion. Compared with the mental health of the community the occasional transplantation of a kidney is not really of great significance.

REFERENCES

HILL, A. BRADFORD (1963) *Brit. med. J.*, **i**, 1043.
MCCANCE, R. A. (1936) *Lancet*, **i**, 823.
MEDICAL RESEARCH COUNCIL (1964) 'Responsibility in investigations on human subjects', reprinted from Report of Medical Research Council for 1962–63 (Cmnd 2382), pp. 21–5.
REPORTS TO MEDICAL RESEARCH COUNCIL (1959) *Brit. med. J.*, **i**, 803; and (1964) *Brit. med. J.*, **ii**, 837.
WOODRUFF, M. F. A. (1964) *Brit. med. J.*, **i**, 1457.

II

Environmental Radioactivity and its Biological Significance

W. V. MAYNEORD

INTRODUCTION

THROUGHOUT recorded time all living organisms have been exposed to small but varying amounts of ionizing radiations. The aim of this review is to sketch very briefly the origins of this natural radiation, its distribution and its variation from place to place, but particularly to consider its possible significance in biology and medicine. We shall find few answers but, I hope, ask some interesting questions.

The materials by which we are surrounded and of which our bodies are made incorporate traces of natural radioactivity. To these we have now added so-called 'artificial' radioactive materials of our own invention and it is one of the ironies of our subject that had these artificial radioactive materials not been associated with the most important political questions of our day, it is doubtful whether anyone would ever have seriously considered radioactivity in a general social context.

In their physical nature and biological action natural and artificial radioactive substances differ but little and are indeed in many instances identical. Man's recent nuclear activities have, for example, added extra carbon 14 to an already existing material stock. It is true that many natural radioactive materials are alpha emitters and that these are relatively rare among artificial radioactive isotopes which mostly emit beta and gamma rays, but there are many beta and gamma emitters among the main natural radioactive 'families', and of course some very important alpha emitters such as plutonium 239 among the artificial materials.

We may, therefore, consider the two sets of radiations together, remembering that the biosphere has been exposed to natural radiation for some three thousand million years, and to the relatively feeble artificial ones for about twenty years.

PHYSICAL PHENOMENA

Following the discovery of the radioactivity of a few heavy elements such as uranium, thorium and radium, it was natural that search should be made to find out if all material objects exhibited the phenomenon. It was soon established that practically all materials examined did indeed show slight traces of radioactivity (Meyer and Schweidler, 1927). The development of modern instrumentation has amply confirmed and extended this view (Rankama, 1954; *Nuclear Radiation in Geophysics*, 1962), the activity in many instances being due to the presence of traces of heavy naturally radioactive metals mixed with a main mass of less active material. Sometimes the activity is due to continuing geophysical processes such as the incorporation throughout the biosphere of carbon 14 produced by the interaction of cosmic rays with nitrogen of the upper atmosphere.

The earth, for our purposes, may be thought of as being some three thousand million years old, so that only original material of a high order of stability can still be present. Uranium 238 and thorium 232 seem to be the parents of most of the natural radioactivity. Their 'daughters' (including radium) form the bulk of the radioactive material to be found in igneous, metamorphic and sedimentary rocks (Meyer and Schweidler, 1927; *Natural Radiation Environment*, 1964). The sole exception of importance is a potassium isotope of very wide occurrence, namely potassium 40, possibly the most important radioactive material in the biosphere and also a main constituent of the lithosphere. It was shown as early as 1907 by N. R. Campbell that all compounds of potassium emitted beta rays, but the activity of wood had been reported by Rutherford in 1903 (Rutherford, 1903), this being so far as I know the earliest recorded radioactivity of a 'biological' material. Potassium 40 contributes some 20 per cent of the dose normally received by our soft tissues, the total radioactivity of our bodies corresponding to several thousand atomic disintegrations per second from this natural element alone. Radio-potassium also constitutes overwhelmingly the major activity of sea-water, the total amount in the oceans being estimated at sixty-three thousand million tons, an activity of the order of a nuclear power station. Since the oceans were the scene of so much evolutionary history and the cradle of life from Cambrian to Silurian periods (perhaps 200 million years) and since so many protozoa and metazoa are small and therefore subject to irradiation from their surroundings, potassium 40 must have been responsible for much of the total irradiation of early living organisms. The alleged similarity of inorganic composition of blood plasma and ocean takes on a new significance.

Many thousands of observations have been made of the radioactive contents of igneous and sedimentary rocks with the result that we are fairly sure of the mean content of such rocks in the accessible regions of the earth's crust (*Natural Radiation Environment*, 1964; *Nuclear Radiation in Geophysics*, 1962). We have good theoretical reasons for thinking that the radioactivity of the earth is largely confined to a layer perhaps 25 to 40 km. thick near its surface. Uranium 238 and potassium 40 are present in the earth's crust to the extent of about three parts per million, while thorium 232 tends to be about four times as abundant on a mass basis. The most abundant radioactive element is rubidium which decays to an isotope of strontium, but its extremely low rate of disintegration and low energy of decay make it of little practical significance.

It should be noted that billions of years ago uranium 235 (the fissile material of some bombs) and potassium 40 were relatively more abundant in consequence of their shorter half-lives. For example, potassium 40 was probably some five times more abundant at the time of formation of the earth's crust than it is now.

Although the amounts of radioactive materials seem so small it was recognized over sixty years ago that they are sufficient to provide major sources of heat in the earth's crust. I have recently estimated (Mayneord, 1964) that the energy liberated by radioactivity beneath the surface of this country is probably more than the electrical power utilized on its surface. We cannot stay to discuss in detail these fascinating geophysical phenomena, but radioactivity may have played an important role in mountain building, in the great convexion currents in the earth's crust which have moved continents, transformed climates and moulded the face of the earth to its present forms. If I had to guess how radioactivity had wrought most havoc among human beings I should say most probably by earth movements, volcanic eruptions and analogous geophysical phenomena.

The radioactivity of soil derived from rocks gives rise to a flux of nuclear radiation from all land surfaces, a flux now known to vary from place to place by orders of magnitude. Igneous and sedimentary rocks vary in radioactive content and hence give rise to varying amounts of radiation above them (United Nations, 1962). In general, the beta-ray dose to humans on the earth's surface is neglected, though it has sometimes to be taken into account as, for example, in considering the allowable contamination of the sea-shore by the effluent from nuclear power stations, or the doses received by native populations living in simple huts with floors made of particularly active soils as in Southern India. The penetrating gamma rays are of more

TABLE 1. Body tissue dose-rates due to external and internal irradiation from natural sources of radiation in 'normal' regions (United Nations, 1962)

Source of irradiation	Dose rate in mrem/y Gonad	Haversian canal	Bone marrow
External irradiation:			
Cosmic rays (including neutrons)	50	50	50
Terrestrial radiation (including air)	50	50	50
Internal irradiation:			
K40	20	15	15
Ra226 and decay products (35 per cent equilibrium)	0·5	5·4	0·6
Ra228 and decay products (equilibrium)	0·8	8·6	1·0
Pb210 and decay products (50 per cent equilibrium)	0·3	3·6	0·4
C14	0·7	1·6	1·6
Rn222 (absorbed in bloodstream)	3	3	3
Total	125	137	122

interest. Table 1 shows (United Nations, 1962) the main components of the ionizing radiations reaching a typical population in the West, whence it will be seen that the terrestrial radiation makes up perhaps one-third of the total, the rest arising largely from cosmic rays and the potassium 40 of our tissues. Table 2 (United Nations, 1962) gives an idea of the average dose rates observed over different types of soil. Typical estimates of the dose-rates over limestone may be 20 milli-rad/year, while over a granite area it may be perhaps 150 milli-rad/year. The mean value for the terrestrial radiation in this country is probably some 50 milli-rad/year Very much higher radiation fields exist in some special districts. For instance, in monazite areas in the States of Rio de Janiero and Esperito Santo, Brazil, a population of 30,000 people is estimated (United Nations, 1962) to live in radiation fields of 500 to 1,000 milli-rad/year. In other volcanic intrusive areas in Brazil peak values of 12,000 milli-rad/year have been recorded. This is 200 times greater than normally observed in this country. A recent report (Roser and Cullen, 1964) notes that higher levels are encountered on some beaches in Brazil. These spots of '*areia preta*', or black sand are, moreover, preferred spots for sun and sand bathing and show gamma-ray levels up to 2 milli-rad/hour (that is, 17,000 milli-rad/year) This, we note, is

TABLE 2. Measured terrestrial gamma dose-rates out of doors in various countries (United Nations, 1962)

Country	Dose-rate in air mrad/y	Comment
Austria	47–56	
France	45–90 180–350	Limestone Granite and shales
Japan	23–37 79–119	Kanto loam Granite areas
Sweden	70–100 60–120 50	Stockholm street Igneous rocks Clay
U.K.	18–61 77–155	Sedimentary rock or clay Granite areas
U.S.A.	45–130	Measurements in 23 States

three times the I.C.R.P. maximum permissible dose-rate level for occupational workers and about 100 times the maximum laid down for the irradiation of general populations. If this beach were part of an atomic energy establishment it would be a controlled area surrounded by radiation warning notices and under the supervision of a responsible protection officer! As in many other parts of the world where exceptionally high radioactivity exists these areas in Brazil are exploited for their alleged curative value and many thousands of Brazilians take their vacations there, coming in contact with the monazite sand and drinking waters of high radioactive content.

In Kerala, India, dose-rates up to 4,000 milli-rad/year have also been reported (United Nations, 1962). We see, therefore, that there are very large differences in the natural gamma rays over the land surfaces of the earth. Moreover, the significance of these differences in dose-rates could be exaggerated by the possibility that food grown in these districts may also be more radioactive. Eisenbud and co-workers (Eisenbud, Petrow, Drew, Roser, Kegel and Cullen, 1964) in Brazil found some high values of radium content in food from radioactive areas, particularly in leafy vegetables such as lettuce and cabbage.

Much more information is required before anything approaching a comprehensive picture of the whole earth's surface can be built up. Oddly enough the radium contents of many drinking waters from these areas of increased radioactivity are not particularly high (Eisenbud *et al.*, 1964) being, for example, less than observed in some American sources (Stehney, 1955; Lucas and Ilcewicz, 1958; Lucas and Krause, 1960; Marinelli, 1958). This agrees with our own findings that in many instances the waters from radioactive areas in this country tend to be low in long-lived activity (Turner, Radley and Mayneord, 1961).

The gamma dose-rate inside buildings is generally different from its value out of doors (Spiers, 1960). It is increased by the radioactive content of the building (bricks were early shown to be active) and decreased by the shielding effects of the walls. Some concretes incorporating shale give unusually high values of dose-rates as do some fibre-board materials used in modern house construction. The relatively high gamma doses inside granite houses (Spiers, 1960) are well known. A small increase in dose-rate (1 to 2 milli-rad/year) may be produced by the accumulation as a result of poor ventilation of natural radioactive gas radon 222 and radon 220 produced respectively by alpha disintegration of two isotopes of radium.

Town and industrial atmospheres are usually more highly charged with radon than those of country districts. It is also interesting that air below the surface in the interstices of soil often contains 1,000 times the concentration of radon found in free air above, and closed spaces such as caves have been known for many years to have high radon content derived from gas released in the decay of radium and diffusing into the limited volume. The radon in soil delivers appreciable doses to the roots of plants and may have been important as a component of the radiation dose given to primitive organisms emerging from the sea on to land.

The estimation of the mean dose to a human population exposed to natural gamma rays is not easy and insufficient data are available to give a reliable estimate of the mean dose-rate to the world's population from these natural environmental gamma rays, but a reasonable guess might be 80 mill-rad/year in air and perhaps 50 milli-rad/year to gonads or bone (United Nations, 1962).

A second major component of irradiation of human beings arises from cosmic rays (United Nations, 1962) concerning which so much has been learned recently by space exploratory vehicles. At sea level we are chiefly irradiated by high energy muons. These rays are strictly outside the scope of this review but we note that they will become of importance when supersonic aircraft fly regularly at perhaps 70,000 feet, particularly during periods of 'solar flares' when large increases of intensity can occur. These phenomena seem unlikely to be of much significance to individual passengers in such aircraft but they may be to crews while the existence of belts of radiation above the earth's surface may limit the possible or probable routes of egress for space-craft. As we have seen, at sea level cosmic rays furnish about one-third of our natural radiation dose.

Having looked at the solid earth and upper atmosphere we naturally ask what is the natural radioactive content of the oceans, lakes and rivers? Again, many thousands of observations have been made (United Nations, 1962; Meyer and Schweidler, 1927; Folsom and Harley, 1957) and the values cover a wide range but generally we might say that the uranium, thorium or radium contents of sea-water may be hundreds of times less than those of soils, but that the activity is variable from ocean to ocean and at different depths. The important activity in sea-water is potassium 40 though the oceans contain perhaps 400 tons of radium (1,000 megacuries).

Nearer to our practical problems are questions of the radioactivity of surface and particularly of drinking waters. It has been known for many years (Meyer and Schweidler, 1927) that natural waters and, in particular, spa waters contained radium and also radon gas in relatively high concentrations (Mayneord, 1964). A few years ago my colleagues and I measured about one hundred drinking waters used in the United Kingdom by fair-sized populations and found very variable amounts of radioactivity in them (Turner, Radley and Mayneord, 1961). Many measurements have also been made in the United States and give a fair overall picture of the situation (Stehney, 1955; Lucas and Ilcewicz, 1958; Lucas and Krause, 1960; Hursh and Gates, 1953). The highest activities in drinking waters are in spa waters. Next in order of decreasing activity come waters from boreholes in geological strata other than chalk, followed by waters from boreholes in chalk and surprisingly last, surface waters from rivers, lakes and reservoirs even though derived from areas of pre-Cambrian formations. The latter often contain little suspended matter and this is the cause of their low activity. If we assume a daily intake of 2·5 litres of water per head of population, the daily intake of long-lived activity from drinking water may vary by a factor of 500 : 1 from place to place and amounts roughly of 0·01 $\mu\mu$c to 60 $\mu\mu$c. The radon intake may easily reach 10,000 $\mu\mu$c.

During recent years a number of authors have studied the natural alpha radioactivity of many hundreds of food samples and find, again, a very wide range of natural radioactivity, largely radium and its products (United Nations, 1962). The ratio of most active to least active foods amounts to 20,000 : 1. In general fruits and vegetables have very low natural activity with meat and milk slightly higher. The most active type of food met with in bulk is cereals (Turner, Radley and Mayneord, 1958*a*; Mayneord, 1960*a* and 1960*b*), but the radioactivity of Brazil nuts is most surprising and forty times higher still. Most of us ingest perhaps 70 per cent of our daily ration of alpha-activity as cereals and related foods (Turner, Radley and Mayneord, 1958*b*). Usually the intake of natural activity in this country from foods is an order of magnitude greater than from waters but this generalization must be treated with caution and does not hold for some American or even British populations (Stehney and Lucas, 1956; Marinelli, 1958).

When we examine the air and its radioactivity we meet much complexity. The radon content of air varies widely from place to place and from day to day. On dull foggy days in cities the content tends to be high while it is usually very low on clear days in the country or at sea. The content of radon over a strongly emanating soil has a well-marked diurnal rhythm being low at midday and high in the early hours of the morning. It varies with barometric pressure, changes of pressure and wind direction.

One of the most interesting developments of the last few years has been the discovery of a natural 'fall-out' cycle (Mayneord and Hill, 1959). Radon diffuses from the surface of the earth into the upper atmosphere at an average rate of about 1 atom per square centimetre of the earth's surface every two seconds, yet this apparently minute leakage has fascinating and observable effects in the biosphere. Following the emergence of radon into the atmosphere the daughter products attach themselves to aerosols and dust particles, continuing to diffuse upwards (Krebs and Stewart, 1962). The lead-bismuth-polonium 214 sequence transforms to a second series of lead-bismuth-polonium 210 products which are washed out in rain and thus brought back to earth (Hill, 1960). The polonium 210 may then be ingested along with grass by sheep and may be found in very appreciable quantities in their kidneys (Hill, 1962). It is scarcely surprising that we find polonium 210 to be the chief alpha-emitting natural radioactive material in human soft tissues (Hill, Osborne and Mayneord, 1964). We estimate that it delivers doses of 2 to 10 milli-rem/year to human gonads (Mayneord and Hill, 1964). This finding is highlighted by the discovery that bones (and we suspect soft tissues) of some Eskimos and Laplanders have an alpha activity due to polonium 210 perhaps fifty times that normally observed in this country. This arises from the continuous deposition of polonium 210 on lichens which are the chief food of reindeer and caribou, which store it with the artificial radioactive material caesium 137 in their soft tissues. These animals are in time eaten by the Northern inhabitants. This soft tissue chain may prove as important genetically as that in bone, which has been studied for so long. Let me, however, point out that normal concentrations signify on the average only a single atom of polonium 210 in perhaps half a million living cells, but that there is evidence that some types of rare atoms tend to be located in particular rare kinds of cells thus providing a possible mechanism of significant action by only a few atoms.

There are many other sources of natural radioactive materials around us (Medical Research Council, 1960; United Nations, 1962). A luminous time-piece may contain perhaps 2 to 3 million μμc of radium, an incandescent gas mantle has an activity such that 1 mg. contains more radioactivity than the usual content of the human body. Automatic lighter flints, powders for polishing and cleaning, even dental fillings, have sometimes relatively high activity. Our few measurements of animal bones have convinced me that man is relatively among the less radioactive species!

Many measurements have been made during the last few years of the natural radioactivity of human tissues (Mayneord, 1964). Radium 226 is normally present in bone to the extent of perhaps 100 μμc per skeleton. Thorium 228 is also present and so are the later members of the radium radioactive series though not in the equilibrium amounts as the gaseous member radon partially escapes into the blood stream and is exhaled.

It seems that our skeletons do not get more active as we grow older and foetal bone shows as high a specific activity as adult bone. This is perhaps not surprising since its materials are derived from adult maternal tissues. We need many more observations on children particularly in the one to five years age group (Mayneord, 1960).

Small but perhaps significant quantities of carbon 14 and tritium make up the rest of our normal radioactive content together with potassium 40 which supplemented by polonium 210 constitutes the main activity of soft tissues. In the surfaces exposed to the outer world, however, as for example the lungs, we find traces of other materials particularly insoluble particles bearing nuclides such as thorium 232, uranium 238, and small quantities of plutonium 239 (Hill, 1962).

As is very well known industrial workers exposed to radium may, of course, show levels of activity perhaps a thousand times normal and may consequently develop pathological changes including bone sarcomas (United Nations, 1962; United Nations, 1964). Their soft tissues show only a rise by a factor of perhaps three (Mayneord, 1960).

We have, therefore, a picture of a naturally radioactive human environment leading to complex cycles of metabolism. At the same time we receive radiation from the ground, the soil, the buildings and ourselves. We are irradiated at all times by an intensity of natural radiation which would cause an ionizing event in each of the cells of our body on an average about once every two or three days. Is this natural radiation and radioactivity of any significance to our health and has it played any important part in biological evolution?

Biological Significance

In a very speculative mood a case could be made that radioactivity and its associated radiation have been of significance in a variety of ways. Of recent years there has been much renewed discussion of the 'origin of life' or, at least, of the origin of materials required by very simple organisms (Oparin, 1964). The original atmosphere of our planet was probably reducing and perhaps consisted of hydrocarbons such as methane, together with ammonia, water vapour and hydrogen sulphide, but no significant amount of oxygen. It was thought that important organic compounds could arise under the action of ultraviolet light or electrical discharges in such an atmosphere. Recently it has been shown experimentally that *inter alia* a few natural amino acids can be produced in such conditions, and since it is difficult to visualize a primeval form of life not centred around proteins these findings open exciting possibilities. Investigation is now tending to be centred on the formation of such biologically important compounds by heat, but the very high ionizing conditions to be found in soils and the well-known significance of ionization in facilitating such reactions inevitably raises the question whether natural radioactivity might not have played a part in the formation of the first bio-compounds. Nuclear energy may well have been involved in the emergence of life from the 'nutrient broth'.

The question might, of course, be asked whether ionizing radiation is necessary to life. It seems improbable for ionizing radiation is *par excellence* an indiscriminate destroyer of biological organization. Attempts to change the fraction of potassium 40 in living organisms and so their internal dose have given negative results but it is interesting to ask whether there are any organisms not naturally exposed (Folsom and Harley, 1957). Where would we expect the lowest natural radiation? Clearly not at high altitudes owing to intense cosmic rays; certainly not below ground unless some less active geological formations are discovered. Some salt mines are stated to have very low natural backgrounds. A man standing above a granite plane surface receives from granite about half the radiation which might strike him if completely surrounded. If he wishes to escape radiation he should hastily take a boat and row out to sea where he will receive only cosmic rays plus half the very low radiation from the ocean. A large fish near the surface of the water may receive perhaps 65 milli-rad/year and at 100 metres deep only about 30 milli-rad/year, almost entirely due to its own potassium 40. A very small micro-organism at a depth of 100 metres may receive less than 5 milli-rad/year in the ocean. This neglects its own radioactivity and some marine plankton contain relatively high concentrations of alpha emitters and so receives high doses (Cherry, 1964). It is, however, in deep fresh water that we should look for the most extreme variation in natural exposure. If an organism is small (a few microns in diameter) and lies in deep clear radioactive free water and stays away from the bottom, it could remain remarkably free from radiation receiving not more than 1 milli-rad/year. It would be interesting to find out how phyto-plankton of 10 μ diameter that seek the deeper portions of the euphotic zone of clear lakes respond to this low dose-rate since only one organism in perhaps 500 would be expected to experience an ionizing event before it divided.

Such considerations lead on to questions as to the significance of ionizing radiation during evolution. The first naïve question might be whether it raised mutation rate and so provided necessary flexibility and speeded evolution. Perhaps more important we may ask whether radiation has widened the possible limits of variation, a question closely related to whether or not the mutations brought about by ionizing radiations are the same as those occurring naturally by other mechanisms, the radiation merely increasing their frequency. I personally have never been convinced that they were the same (Mayneord, 1964) and the study of 'Doubling dose' (Medical Research Council, 1960), the known unspecific character of ionization in general, studies of bacteria and the known selective character of chemical mutagens suggest that the frequency distribution of hereditary defects resulting from a specific increase in the level of exposure to radiation might not parallel the existing natural spectrum which is itself variable. This is not the place to consider in detail the possible effects of natural radiation during the early stages of terrestrial evolution for we know very little of the changes of natural radiation over hundreds of millions of years (Mayneord, 1964). The radiation poured into the germ plasm of some long-surviving phyla is very large and one wonders how unicellular organisms, multiplying by binary fission, avoid transmitting defects in cytoplasmic organization as well as genetic code from parent to progeny.

MEDICAL ASPECTS

To cease these interesting but possibly idle speculations, we now ask is there any evidence that natural radiation has influenced the incidence of any disease known to be produced by ionizing radiation at higher dose levels, as for example leukaemia or bone sarcoma? Is there any evidence of genetic effects in areas of high activity?

It has been suggested *inter alia* that all ageing is due to natural radiation background. My only comment would be that I know of no evidence and it seems unlikely. We have seen that certain localized areas support sizeable populations with success, yet have a natural background up by a factor of 10 or even 100 on that observed in this country. On the other hand the Medical Research Council in its first report (Medical Research Council, 1956) considered it 'unlikely that any authoritative recommendation will name a figure for permissible radiation dose to the whole population, additional to that received from natural background, which is more than twice that of the general value for natural background radiation'. That is, three times the present levels in this country is to be regarded as at least undesirable. So far, little or no conclusive evidence has been obtained from the study of naturally irradiated human populations. The determination of the mean dose received by a scattered population is difficult owing to the very variable radiation levels encountered over small distances, the varying fractions of time spent in and out of doors by different persons and because of movement within the area. If we are thinking of genetic effects these physical difficulties are, however, small compared to those arising from the complexities of social and thence genetic structure, race, religion, age structure of population, mating habits (as for example probability of matings between close relatives). Although the genetic effect on a population is defined in the ideal by the product of mean dose and number of individuals, in practice the situation is more complex. Mendelian populations are rarely sharply defined and differences in language, religion, economic status, education and profession, help to subdivide a large population into smaller and smaller groups within which random mating plays a part and becomes more probable. Love is blind and can often stumble only to the limits of the village or the social club. Add to this our ignorance of the rate at which increase of radiation dose increases the probability of a given mutation in man, coupled with the long periods of many generations required for the build-up of some genetic equilibrium, all of this in a shifting social panorama probably causing genetic change in many other unknown ways. Change of mean age of marriage for instance or improved transport facility may be important. We see why we are unlikely to obtain information quickly. Moreover, we are looking for very small changes and these require the study of very large populations and thence the necessity for the inclusion of a wide range of physical and social conditions. We must, I think, conclude that so far no one has demonstrated any genetic effects in man attributable to varying natural or artificial background; perhaps we should not have expected them to, though the search goes on. In a report published by the World Health Organization (W.H.O., 1959) detailed study was made of the type of information needed, the statistical problems involved and the many other factors to be taken into account in such studies.

With respect to leukaemia, it has been thought from several lines of argument (such as the study of atomic bomb survivors; of irradiated patients suffering from ankylosing spondylitis; of the life shortening of radiologists; of irradiation of children for thymic enlargement; of leukaemia in American radiologists; of leukaemia in children irradiated *in utero*) that the probability of inducing leukaemia per individual per rad is 1 to 2 per million. A recent report of UNSCEAR (United Nations, 1964) on radiation carcinogenesis in man regards this as an upper limit for effects at very low doses but considers that the risk of leukaemia per unit dose might be several times higher in children irradiated *in utero* than in adults. Taking the figures at their face value, if the whole-body dose per person in a lifetime due to natural radiation is some 10 rad, this implies that some 10 per cent to 20 per cent of observed leukaemia might be due to radiation background.

However, there is now evidence against the simple assumptions (Lewis, 1957; Loutit, 1962), that there is a linear relation of dose to leukaemic effect irrespective of time since irradiation, age at irradiation or dose-rate. There is, again, the necessity to study very large populations if precision is to be obtained. For example, if the annual natural occurrence of leukaemia is of the order of 50 cases per million persons and 10 per cent is caused by radiation, then if the annual radiation were increased by a factor of two, the incidence would only rise from 50 to 55 cases per million. To have satisfactory confidence that such an increase was real one would have to study a population of several millions. The reported number of leukaemia cases is increasing by about 2 per cent to 7 per cent each year in most countries, and against this varying background it is unwise to draw any conclusions as to the possible correlation of leukaemia induction with normal low level exposure. There are, moreover, many complex questions involved in the use of control populations, for example unirradiated spondylitics. Possibilities of error arise from an association of leukaemia and rheumatic disease or the use of chemotherapeutic agents such as antipyrene, gold salts or phenylbutazone known to depress bone marrow activity. Bean (1960) for example reported the development of leukaemia in six non-irradiated patients having received phenylbutazone.

Court-Brown and Spiers (1960) made a survey of leukaemia in Scotland since parts of Scotland, notably Aberdeenshire and the city of Aberdeen, are rich in granite as noted above. Dose measurements indicated that the average doses in bone marrow was 80, 86 and 101 milli-rad per year in Edinburgh, Dundee and Aberdeen respectively. If natural background is responsible for 10 per cent to 20 per cent of natural leukaemia one would not expect a 20 per cent difference to show in a city of less than 200,000 people like Aberdeen. In fact, over an 18 year period the death rate from leukaemia was nearly half as much again as the natural Scottish average. Either we have under-estimated the effect of background by a factor of 100 or other factors are far more important (Court-Brown and Spiers, 1960; Court-Brown and Doll, 1960). I certainly incline to the latter view.

Attempts to prove higher incidence of leukaemia in Cornwall, where there are high backgrounds, have also proved inconclusive (Turner, 1962). Craig and Seidman (1961) have reported that the increase of leukaemia in the one hundred and sixty-three metropolitan areas of the United States bears no demonstrable relationship to differences in the amounts of cosmic radiations received. On the other hand Gianferrari (Gianferrari, L., Serra, A., Morganti, G. *et al.*, 1962) made a survey of births, deaths and other relevant variables in some communities of the Cervo Valley in Italy, where the background radiation is higher than normal (average 15 r in 30 years) and in a near-by area, similar geographically, socially and economically, except for one-fifth (3 r in 30 years) of background radiation. In the high radiation area the observed proportion of deaths from cancer was higher than expected in every age group but the increase was statistically significant only in the sixty to eighty age group. These findings raise again the question whether radiation raises the incidence of a given condition by a fixed amount per unit dose or raises an existing incidence by a given factor.

I have been unable to obtain significant data in relation to regions of high activity around thermal springs as, for example Bad Gastein, Austria. With a population of only 5,000 in this spa this is not surprising. It is, however, maintained by Hittmair (1957) that in the active areas cancer is a less common cause of death, being some 6·1 per cent of the deaths during a 10 year period (1947–56). Clearly such data require detailed knowledge of population structure and many other conditions before they can be interpreted.

A survey of the mortality from congenital malformations in different geological regions of the United States (Kratchman and Grahn, 1959) has suggested that higher incidences are associated with geographical areas that contain major uranium ore deposits, uraniferous waters or helium concentrates. Another survey by Wesley (1960) suggested correlation of congenital malformations with geomagnetic latitude and hence cosmic rays but neither of these surveys is convincing. There is no certain evidence that natural radiation influences the cancer rate in man, and this is not surprising since its intensity is rarely 1 per cent of the lowest dose-rate so far shown to induce malignant disease in man or animals.

The external environmental range of dose-rates of populated areas of the United States stretches from 70 to 175 milli-rad/year, with lower dose-rates prevailing in the more populated Eastern to Mid-Western States (Solan, Lowder, Zila, LeVine, Blatz and Eisenbud, 1958; Bugher and Mead, 1958). An examination of the frequency of bone sarcoma over the same area has been made (Bugher and Mead, 1958) in relation to this low level exposure. Primary bone tumours were estimated to cause 2,000 deaths a year in this area, decreasing slightly from 1949–55. There is a peak of incidence at about 20 years of age. The analysis of mortality data for bone tumours over a period of seven years involving 14,000 deaths in $1{\cdot}6 \times 10^9$ man years, failed to show any increase in the frequency of the diseases in the regions of highest radiation exposure. Indeed, the contrary situation appeared to prevail in accord with earlier observations by Auerbach (1958) based on the years 1949–51.

The high radium 226 content of certain drinking waters in Illinois is being studied by the United States Public Health Service (Samuels, 1964) in connexion with the incidence of bone cancer in this area as well as in relation to deaths from other causes. 3 pc/litre is the suggested limit of radium 226 content of drinking water but some 50,000 people are exposed to over 10 pica-curies of radium 226 per litre. A similar state of affairs holds in certain regions of this country (Turner, Radley and Mayneord, 1961). The Harvard School of Public Health is also conducting studies of the incidence of leukaemia, malignant neoplasms of bone and congenital malformations in the populations of selected areas of Maine, New Hampshire and Vermont, having high natural backgrounds. It is considered doubtful that any significant correlation between radiation exposure and incidence of biological effect will be found owing to the limited size of population and the relatively small range of radiation exposure expected, but it was thought that such a venture would provide considerable methodological information as well as experience useful in further studies of a similar nature. It should be mentioned that the study of the incidence of cancer and congenital malformations in northern New England shows up to the moment no indication of any correlation with bedrock radioactivity estimates (Lowder, Segall and Condon, 1964).

A project for studying the geographic distribution of cancer within a single country as related to environment factors has been described by Lawrence and Chen (1959), while some years ago Myers and Hess (1952) used a direct method of measurement in attempts to correlate biological and physical data. Similarly, Gentry, Parkhurst and Bulin (1959) used geological guides to estimate background radiation of houses associated with congenital malformations in New York State residents, but such methods have many pitfalls (Pinkerton, Chen, Hutchins and Schrohenloher, 1964).

In Great Britain, one of the few investigations carried out in this field was by my colleague Turner (1962) who, following our investigations of radioactivity of food and water and the corresponding variations of daily intake of radium 226, investigated the correlation between the incidence of a number of sites of cancer and the radioactivity of drinking water. The data do not provide any evidence of any association between cancer mortality rates (i.e. carcinoma of the breast, uterus, trachea, lung, bronchus, stomach or leukaemia) and the natural radioactivity of the drinking water. Indeed, the areas of Wales which show surplus mortality for gastric cancer possess drinking waters with low content of activity. A similar negative correlation is suggested for primary bone tumours.

To sum up, I personally must conclude that in spite of very large numbers of observations there is, as yet, no direct evidence that the geography of life and death has any correlation with natural radiation background. The intricacies of the radiation pattern, together with the complexity of the shifting social picture make the certain demonstration of effect at natural levels of irradiation extremely difficult.

It is necessary to study large populations not so much to obtain conditions of formal statistical significance, but rather to observe even a very small number of actual cases of the condition being investigated. Expressed in a different way the difficulty lies in the extremely different orders of magnitude of observations of the physical and pathological phenomena. To give a single example. A physicist may detect a single ion pair, certainly a single alpha particle. In the skeletons of one million members of the population of this country some 10^{14} alpha particles are emitted per year setting free perhaps 10^{19} ion pairs. The number of deaths per annum due to primary bone tumours in the same number of people is perhaps five. It is this fantastic but fortunate *im*probability that an observable physical event will give rise to the pathological entity which raises our naïve hopes of correlation but in the end dashes them to the ground. I sometimes wonder, too, if the word 'cause' is legitimately applied to such a tenuous connexion.

REFERENCES

AUERBACH, H. (1955) Argonne National Laboratory Report No. 5426.

BEAN, R. H. D. (1960) *Brit. med. J.*, **ii**, 1552.

BUGHER, J. C. and MEAD, P. A. (1958) Second International Conference on Peaceful Uses of Atomic Energy. United Nations, **23**, 165.

CHERRY, R. D. (1964) *Nature*, **203**, 139.

COURT-BROWN, W. M. and SPIERS, F. W. (1960) *Brit. med. J.*, **i**, 1753.

COURT-BROWN, W. M. and DOLL, R. (1960) *Proc. Roy. Soc. Med.*, **53**, 762.

CRAIG, L. and SEIDMAN, H. (1961) *Blood*, **17**, 319.

EISENBUD, M., PETROW, H., DREW, R. T., ROSER, F. X., KEGEL, G. and CULLEN, T. L. (1964) *Natural Radiation Environment*, University of Chicago Press.

FOLSOM, T. R. and HARLEY, J. H. (1957) *Effects of Atomic Radiation in oceanography and fisheries*, p. 28. Nat. Acad. Sci. Nat. Res. Council Publication 551. Washington, D.C.

GENTRY, J. T., PARKHURST, E. and BULIN, G. V. (1959) *Amer. J. pub. Health*, **49**, 497.

GIANFERRARI, L., SERRA, A., MORGANTI, G. *et al.* (1962) *Bull. World Health Organization*, **26**, 696.

HILL, C. R. (1960) *Nature*, **187**, 211.

HILL, C. R. (1962) *Health Physics*, **8**, 17.

HILL, C. R., OSBORNE, R. V. and MAYNEORD, W. V. (1964) *Natural Radiation Environment*, p. 395. University of Chicago Press.

HITTMAIR, A. (1957) *Medizinische Klinik*, **17**, 712.

HURSH, J. B. and GATES, A. A. (1953) University of Rochester Report U.R. 257.

KREBS, A. T. and STEWART, N. G. (1962) *Nuclear Radiation in Geophysics*, p. 241. Springer-Verlag, Berlin.

KRATCHMAN, J. and GRAHN, D. (1959) United States Atomic Energy Commission, Washington, Report TID 8204.

LAWRENCE, P. A. and CHEN, W. Y. (1959) *Amer. J. pub. Health*, **49**, 668.

LEWIS, E. B. (1957) *Science*, **125**, 965.

LOUTIT, J. F. (1962) *Irradiation of Mice and Men*. University of Chicago Press.

LOWDER, W. M., SEGALL, A. and CONDON, W. J. (1964) *Natural Radiation Environment*, p. 907.

LUCAS, H. F. and ILCEWICZ, F. H. (1958) *J. Amer. Waterworks Assoc.*, **50**, 1523.

LUCAS, H. F. and KRAUSE, D. P. (1960) *Radiology*, **74**, 114.

MARINELLI, L. D. (1958) *Amer. J. Roentgenol.*, **80**, 729.

MAYNEORD, W. V. (1960*a*) Appendix E, p. 73. *Hazards to Man of Nuclear and Allied Radiations*. Second Report to the Medical Research Council. Cmnd. 1225. H.M.S.O., London.

MAYNEORD, W. V. (1960*b*) *Clinical Radiology*, **11**, 2.

MAYNEORD, W. V. (1964) *Radiation and Health.* Nuffield Provincial Hospitals Trust, London.

MAYNEORD, W. V. and HILL, C. R. (1959) *Nature*, **184**, 667.

MAYNEORD, W. V. and HILL, C. R. (1964) Heidelberg Symposium on Assessment of Radioactive Body Burden in Man. International Atomic Energy Agency, Vienna.

MEDICAL RESEARCH COUNCIL (1956) *Hazards to Men of Nuclear and Allied Radiations*. Cmd. 9780. H.M.S.O., London.

MEDICAL RESEARCH COUNCIL (1960) *Hazards to Man of Nuclear and Allied Radiations*. Second Report. Cmnd. 1225. H.M.S.O., London.

MEYER, STEFAN and SCHWEIDLER, EGON. (1927) *Radioaktivität.* B. G. Teubner, Berlin.

MYERS, J. and HESS, V. F. (1952) *N.Y. State J. Med.*, **52**, 463.

Natural Radiation Environment (1964) Edited by John A. S. Adams and Wayne M. Lowder. Rice University Semicentennial Series. University of Chicago Press.

Nuclear Radiation in Geophysics (1962) Edited by H. Israel and A. T. Krebs. Springer-Verlag, Berlin.

OPARIN, ALEXANDER I. (1964) *The Chemical Origin of Life.* Charles C. Thomas. Springfield, Illinois.

PINKERTON, C., CHEN, W. Y., HUTCHINS, R. G. and SCHROHENLOHER, R. F. (1964) *Natural Radiation Environment*, p. 919. University of Chicago Press.

RANKAMA, KALERVO (1954) *Isotope Geology.* Pergamon Press, London.

ROSER, F. X. and CULLEN, T. L. (1964) *Natural Radiation Environment*, p. 825. University of Chicago Press.

RUTHERFORD, E. (1903) *Nature*, **67**, 511.

SAMUELS, LARRY L. (1964) *Natural Radiation Environment*, p. 329. University of Chicago Press.

SOLAN, L. R., LOWDER, W. M., ZILA, A. V., LEVINE, H. D., BLATZ, H. and EISENBUD, M. (1958) Second International Conference on Peaceful Uses of Atomic Energy. United Nations, **23**, 159.

SPIERS, F. W. (1960) Appendix D, p. 66. *Hazards to Man of Nuclear and Allied Radiations.* Second Report to the Medical Research Council. Cmnd. 1225. H.M.S.O., London.

STEHNEY, A. F. (1955) *Acta radiologica*, **43**, 43.

STEHNEY, A. F. and LUCAS, H. F. (1956) First International Conference on Peaceful Uses of Atomic Energy. United Nations, **11**, 49.

TURNER, R. C. (1962) *Brit. J. Cancer*, **16**, 27.

TURNER, R. C., RADLEY, J. M. and MAYNEORD, W. V. (1958*a*) *Brit. J. Radiol.*, **31**, 397.

TURNER, R. C., RADLEY, J. M. and MAYNEORD, W. V. (1958*b*) *Health Physics*, **1**, 268.

TURNER, R. C., RADLEY, J. M. and MAYNEORD, W. V. (1961) *Nature*, **189**, 348.

UNITED NATIONS SCIENTIFIC COMMITTEE ON THE EFFECTS OF ATOMIC RADIATION (1962) REPORT. New York.

United Nations Scientific Committee on the Effects of Atomic Radiation (1964) Report. New York.

Wesley, J. P. (1960) *Internat. J. rad. Biol.*, **2**, 97.

World Health Organization Technical Report (1959). No. 166. World Health Organization, Geneva.

III

Ultrasonics in Clinical Diagnosis

P. N. T. WELLS

As long ago as 1920, Langevin produced an ultrasonic depth-sounder. Like many important inventions, Langevin's instrument was simply a new application of older discoveries. In the early part of the last decade, the ultrasonic pulse-echo method, basically Langevin's technique, was first used in clinical diagnosis. Outstanding in this work was that of Leksell (1955), with his successful measurement of the position of the mid-line of the brain through the intact skull.

Basic Principles

Ultrasound is reflected when it strikes an interface between two media of different acoustic properties, and, because it has a finite propagation velocity, this effect may be used for the measurement of distance. Fig. 1 shows a block diagram of the simplest type of diagnostic machine based on the echo principle. The transducer is a piezo-electric disc resonator, which is made to emit pulses of ultrasound by the transmitter. The time-base is started at the same instant as the pulse of ultrasound is emitted into material under investigation. When the ultrasound strikes an echo-producing interface, such as that between media A and B in the diagram, a reflected pulse returns towards the transducer, which produces a voltage proportional to the amplitude of the echo. This voltage is amplified and applied to the y-plates of the cathode ray tube. Thus, two vertical deflexions of the trace occur: the first is caused by the transmitter exciting the transducer, and the second, by the echo. The distance between the two deflexions is equal to the distance between the transducer and the echo-producing interface, multiplied by twice the ratio of the time-base velocity to the acoustic velocity in medium A.

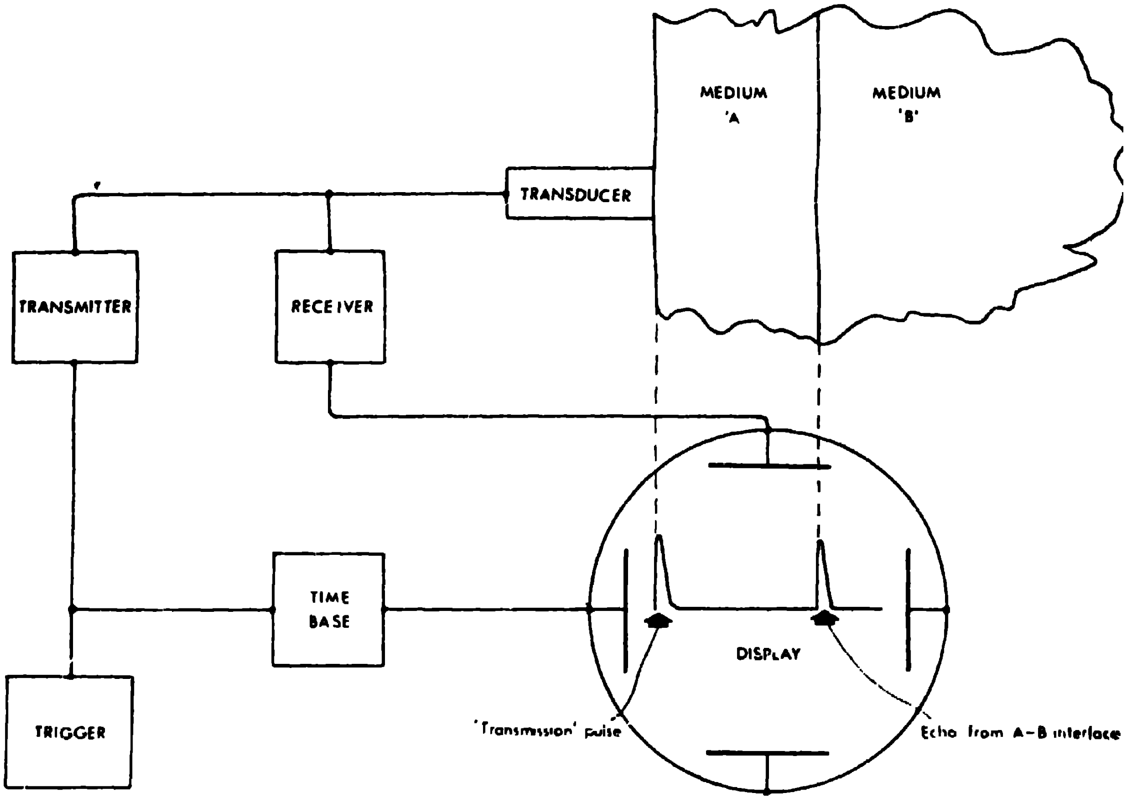


FIG. 1. Block diagram of basic A-scope system.

The trace may be repeated at any desired rate, provided that the ultrasound used to produce one trace has decayed sufficiently before the next is initiated. This simple principle also applies when there are many interfaces producing echoes, as in most biological systems.

FACTORS CONTROLLING PERFORMANCE

(i) *Specific acoustic impedance*

The amplitude fraction R of the echo from an interface depends upon the ratio of the specific acoustic impedances r_1 and r_2 of the media on each side of the boundary, according to the equation:

$$R = \frac{r_1 - r_2}{r_1 + r_2}$$

Specific acoustic impedance is equal to the product of density and acoustic velocity. The specific acoustic impedance ratio for an interface between a liquid or a solid and a gas is very great. As an example, the amplitude reflexion coefficient for a water to air interface is about 0·9995. Thus, the physicist is able to obtain a nearly perfect reflector; but the presence of gas in the path of the clinician's beam effectively prevents any further penetration.

(ii) *Beam shape*

The energy distribution within the beam of a uniform, cophasally vibrating piston may be predicted by interference analysis. The beam may be considered to be split into two zones. In the zone near to the transducer, there are a number of axial maxima and minima, the spacing being closest nearest to the source, and most of the energy is contained within a cylinder. Energy loss in the near zone is due chiefly to absorption. In the far zone, beyond the last maximum, the beam diverges as a cone, and the intensity is reduced with distance both by the inverse square law and by absorption.

In the case of a pulsed ultrasonic beam, such as is used in medical diagnosis, the theoretical prediction of the energy distribution is very complicated. The distribution changes during the beam initiation. For the first half-cycle, the beam is substantially free from diffraction effects; but by the sixth half-cycle, the distribution is close to that for continuous waves. The matter has been discussed in detail by Christie (1962). Thus, the distribution which is observed experimentally for pulsed beams is actually either a time-average of the distribution for the complete pulse, or applies for only part of the pulse.

Conventional schlieren techniques are not applicable to low-power pulsed beams, because the light 'noise' swamps the optical signal. However, the distribution can be measured by plotting the loci of iso-echo amplitudes from a small spherical target. A sphere is used as the target, because it presents a similar surface to the beam, independent of its position. The size of the sphere determines the maximum amplitude gradient which may be observed. A typical result is shown in Fig. 2. In this particular case, the near zone was shorter by some 15 per cent than that calculated for the ideal continuous-wave case.

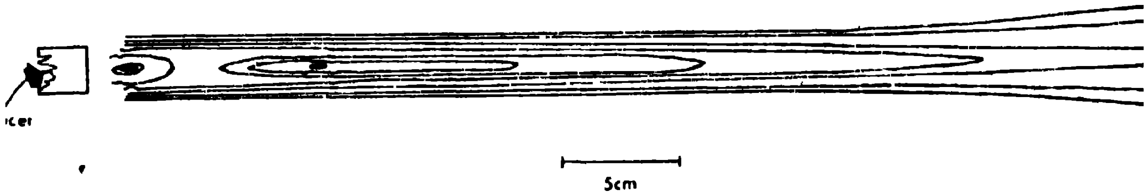


FIG. 2. Iso-echo amplitude curves in water. Plotted for a 1·7 Mc/s, 2·0 cm diameter transducer, using a 0·32 cm diameter steel sphere, at 18° C. 5 dBl. spacing.

The lateral resolution of an ultrasonic system is limited by the energy distribution of the beam. If the beam is wide, good lateral resolution can only be achieved by limiting the dynamic range of the display system. Kossoff, Robinson, Liu and Garrett (1964) have shown that improved lateral resolution can be achieved by the use of a lens or curved transducer arranged to focus the beam. Focusing also has the desirable effect of reducing the decrease of sensitivity with range.

(iii) *Pulse shape*

The depth resolution of a diagnostic system is limited by the rise-time of the pulse presented to the display system. In order to achieve as rapid a rise-time as possible, the bandwidth of the system must be made as wide as possible. Usually, the limiting component is the transducer. A piezo-electric disc which is air-backed has a narrow bandwidth. For the production of continuous waves, this is desirable, and the transducer efficiency can be as high as 40 per cent working into a water load. However, the use of such a transducer for diagnostic work is quite impracticable. Plate I, Fig. 3 shows the echoes obtained from a thin sheet of perspex in water.[1] The upper trace was obtained using a properly designed, well-damped transducer of wide band, and the echoes originating from each side of the perspex sheet are clearly shown; whereas the lower trace was obtained using a narrow bandwidth transducer designed for continuous wave work, and the pulse packet is long and confused.

---

[1] The plates referred to in this paper will be found between pages 50–1.

Wide-band transducers are usually made by fixing a highly absorbent backing to the transducer disc. This has the effect of preventing resonance from building up between the two faces of the transducer. The backing material also reduces the efficiency of the transducer, and this limits the maximum bandwidth which can be achieved in practice.

The higher the frequency of the ultrasound, the shorter the pulse which contains a given number of complete oscillations. However, there is a practical limit to the frequency which can be used in any application: this is because, in any given medium, the absorption increases with the frequency. The echo signals must be an order of magnitude greater than the receiver noise in order to be detected satisfactorily. In order to achieve a good signal to noise ratio, it is often necessary to limit the bandwidth of the receiver by making use of tuned circuits.

The effect of the medium under investigation on the rise-time of the echo is of special importance. The absorption increases with the frequency, and so the higher frequency losses increase more rapidly than the lower frequency losses as the range is increased. Therefore, the rise-time increases with the range. Published figures for absorption in biological materials show a wide divergence, but typical measurements at 1 Mc/s are: blood, 0·09 dB. per cm; fat, 0·68 dB. per cm; muscle, 1·3 dB. per cm; and skull-bone, 4 dB. per cm. In practice, it is not uncommon for the weakest echoes to have an amplitude of 30 or 40 dB. below that of the strongest echoes, and this dynamic range increases with the frequency of the ultrasound. Therefore, the echo rise-time is limited by the material under investigation, either by the limitation of the maximum frequency which can be employed in order to obtain echoes an order of magnitude greater than the receiver noise, or by the limitation of the upper frequency response of the system. This in turn limits the resolution which it is possible to achieve in any particular circumstances.

(iv) *Angulation*

The echo amplitude also depends on the angle of incidence between the beam and the echo-producing interface. This is not important when the echo-producing object is small, and particularly when it is spherical; but when the interface is smooth and extensive, the echo which is received by the transducer is critically dependent upon the angulation of the beam. Fig. 4 shows the effect of angulation on the echo amplitude received

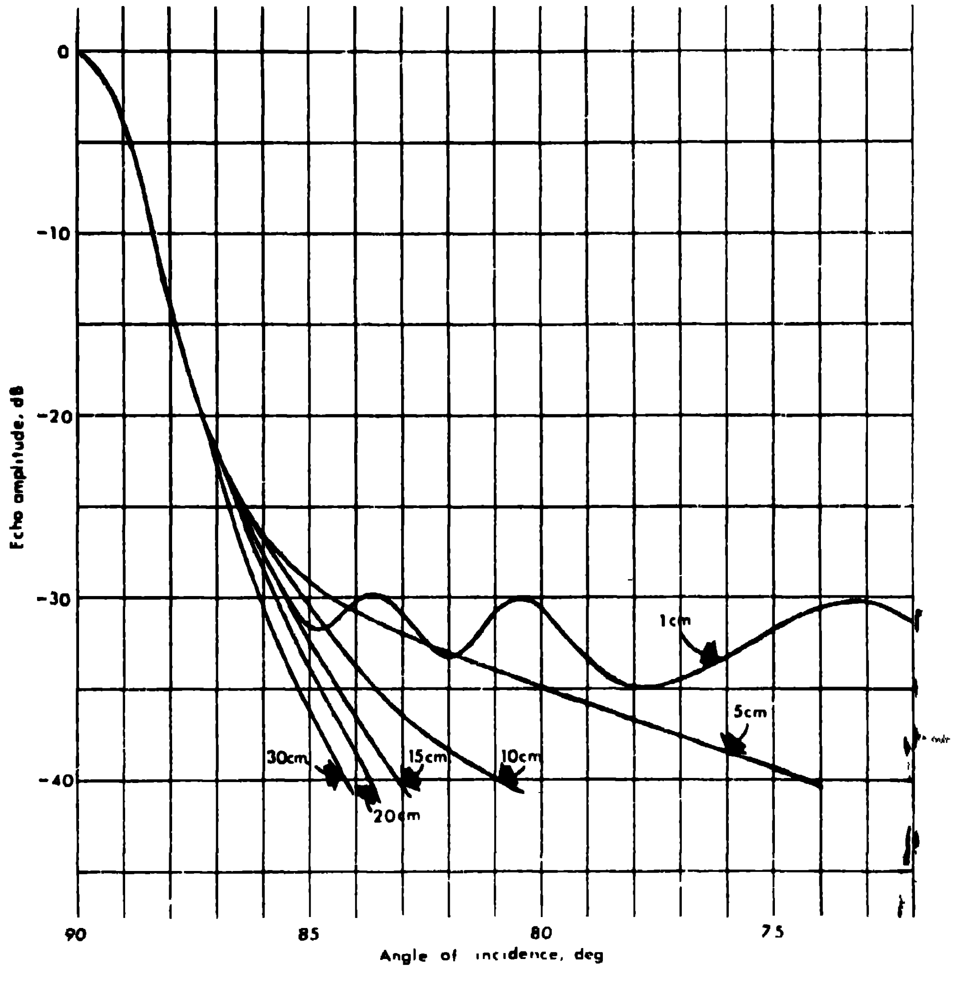


Fig. 4. Effect of angulation on echo amplitude at various ranges. Data obtained for a water to perspex interface, using a 1·7 Mc/s, 2·0 cm diameter transducer, at 18° C. Amplitudes expressed relative to 0 dB. at 90° incidence.

from a water to perspex interface at various ranges. The amplitudes are expressed in decibels relative to the amplitude with normal incidence. Within 3° of normal, the amplitude is reduced by about 22 dB., irrespective of the range. With greater angulation, the subsequent reduction increases with the range. The maxima and minima which occur when the range is small are due to the peculiar phase relationships which exist. They are not important in practice, because the transmission pulse, an artifact chiefly due to the blocking of the receiver by the large voltage from the transmitter, usually extends beyond this region and so makes information retrieval impossible.

(v) *Range*

When the echo-producing interface is very close to the transducer, the echo amplitude exhibits a number of maxima and minima with altering range, similar to those produced by changing angulation. However, in clinical diagnosis, this effect is again unimportant, because the transmission pulse usually extends for a few centimetres in front of the transducer. Beyond this region the echo amplitude falls off smoothly with increasing range, due both to absorption and to the inverse square effect. In clinical practice the attenuation of the echo amplitude with range is often quite large. Echoscopes designed for medical use have a special receiver, the gain of which is automatically varied with time so that similar interfaces produce display deflexions of equal amplitude irrespective of the depth of origin. In biological systems, there are usually a number of interfaces spaced in depth along the ultrasonic beam. Each of these interfaces reflects a fraction of the ultrasound, and so reduces the amount of energy which is transmitted. This happens not only as the pulse is travelling away from the transducer, but also as the echo returns, and it causes a further reduction in echo amplitude.

(vi) *Resonance artifacts*

One of the most troublesome artifacts which is experienced in diagnostic systems is caused by strong echoes, which act as sources of secondary echoes. Usually, the echo bounces backwards and forwards between the transducer and the echo-producing interface, and each complete cycle of this resonance causes an echo pulse to originate at a new position on the time-base. Of course, the amplitudes of these echoes are reduced in the usual way as the effective path length increases, and in accordance with the reflexion coefficients at the boundaries, and so they eventually fall below the sensitivity of the receiver. However, it is not necessary for the ultrasonic path to be linear, and any echo which suffers multiple reflexions before it returns to the transducer will cause a similar artifact at the appropriate point on the time-base.

(vii) *Suppression*

Often, an echo does not produce a single spike deflexion if it is displayed after straightforward amplification and rectification, but gives a rather jagged waveform. This is troublesome when two echoes occur close together on the time-base, because it is then difficult to decide where one echo ends and the next begins.

If the performance of the amplifier is modified so that only signals above a certain amplitude are displayed, the echogram quality is improved. It is necessary to increase the amplification, so that the signals which remain after the suppression process may be large enough to produce useful deflexions.

The higher amplification leads to reduced bandwidth. The display deflexion occurs close to the time of maximum echo amplitude, which does not correspond to the leading edge of the pulse. The dynamic range of the display is reduced, so that only signals of nearly equal amplitude can appear simultaneously if serious overloading is to be avoided.

INTERPRETATION OF CLINICAL ECHOGRAMS

As an example, Plate II, Fig. 5 shows a normal cranial echogram, obtained by the usual technique from a healthy subject. The upper trace was taken with the transducer on the right temporal region, and the lower, on the left. The first deflexions are collectively called the transmission pulse. There is then a number of small echoes, the origin of which is uncertain, followed by the well-known 'mid-line' echo; then there are further unidentified echoes, and finally the echoes from the inside and the outside of the skull. The mid-line echo appears in the same position along the time-base in both tracings, which means that the mid-line structures were not displaced.

Distances measured along the linear time-base are not in direct proportion to the distances between the interfaces which produced the corresponding echoes along the ultrasonic beam.

This is because the velocity varies from about 1,600 m.sec.$^{-1}$ in soft tissue to about 4,800 m.sec.$^{-1}$ in hard bone.

(i) *Limitations*

A certain amount of skill in the manipulation both of the transducer and of the instrument controls is required in order to obtain a satisfactory echogram. There must be no air trapped in the hair between the transducer and the scalp: this requires liberal application of a coupling liquid such as olive oil. The beam must be directed across the skull symmetrically. The angulation must be such that echoes return to the transducer both from the mid-line structures and from the far side of the skull. Even in the healthy subject, some compromise is necessary. Refraction occurs because the beam does not strike every interface normally; thus, the path of the beam does not lie along a straight line. In cases with displaced mid-lines, it may be impossible to obtain a satisfactory echogram in the conventional sense, although the experienced clinician can sometimes use this very difficulty to aid his diagnosis.

(ii) *Clinical applications*

The limitations imposed on the technique by the restriction of useful angulation make possible a very elegant diagnostic procedure. Only when the ultrasonic beam passes symmetrically through the skull does a significant echo arrive from both sides. This is the basis of the method developed by Willocks (1962) for the measurement of the bi-parietal diameter of the foetal head from outside the maternal abdomen.

It is sometimes desirable to obtain useful echoes from very close to the surface of the patient; as an example, this is a requirement for the measurement of the lengths of the various components of the eye. This is impossible with a conventional transducer, because of the length of the transmission pulse. The difficulty can be overcome by the use of a water-filled tube in front of the transducer, so that the transducer stands off by a distance greater than the length of the transmission pulse.

There are many other applications of the A-scope method in clinical diagnosis, which are described in detail in the excellent reviews of Gordon (1963) and Newell (1963).

ULTRASONIC TOMOGRAPHY

The interpretation of an A-scope display is particularly difficult where there are many interfaces at many angles. However, the situation is much improved if the type of display system originated by Howry (1955) is used.

(i) *Basic principles*

The display is made possible by employing brightness instead of amplitude modulation. A dot of light appears on the time-base for each echo above a certain amplitude. This type of display is called a B-scope. The ultrasonic beam is scanned across the part of the patient to be examined, and the direction of the time-base across the display is coupled to the axis of the transducer, so that a tomograph is built up on the display by a series of dots. The picture can be integrated photographically or on an electronic storage tube. The information contained in such a tomograph is the same as would be contained in an equivalent number of simple A-scan echograms, but the relationships between the echoes can be more clearly perceived. In addition, many artifacts, particularly those due to noise, occur as isolated events, whereas echoes due to structures are cumulative.

(ii) *Water-bath coupling*

Early B-scope techniques employed a linear transducer motion across a water bath, with the patient actually immersed. Apart from the inconvenience to the patient, there are a number of difficulties associated with this type of transducer movement. In all water-bath systems, the body wall is a source of resonance artifacts, and so it is essential for the transducer to stand off from the patient by a distance equal to the required penetration. Because the beam does not necessarily strike the body wall normally, refraction occurs unless the temperature of the bath is controlled so that the velocity in water is equal to that in the patient. Finally, the method does not make any allowance for the effect of angulation, and many echoes are not received. The angulation difficulty can be greatly reduced by the employment of a more complex transducer movement, such as the compound circular scan. In this system, the transducer is moved around a circular path, with its axis executing a small oscillation about the radius of the circle. However, even this type of scan is not without its limitations, and when examining structures curved in all directions, many scans in different planes are necessary in order to be reasonably sure that all the interfaces have been found.

Water-bath coupling of the transducer permits a fixed transducer path, and so accurate positional information may be easily obtained. Acoustic coupling is good, even if a flexible membrane is used between the water-bath and the patient to overcome the immersion problem. However, the machine is immobile, which means that it cannot be brought to the bedside of very ill patients.

(iii) *Direct-contact coupling*

A more convenient machine from the clinician's point of view is one which employs a transducer which actually moves in contact with the part of the patient to be examined. Such an instrument has been built by Donald and Brown (1961), and has been used with great success, particularly in gynaecological cases. Of course, the relative freedom of transducer movement which is required makes accurate positional information much more difficult to obtain. Lack of contact makes the method difficult to use on some sites. Greatorex and Ireland (1964) have developed a contact scanner especially for the brain, and they have found coupling to be a major problem.

(iv) *Intravenous transducers*

One of the most recent developments in ultrasonic tomography makes use of intravenous transducers. The method is due to Kimoto, Omoto, Tsunemoto, Muroi, Atsumi and Uchida (1964). The transducer used in these investigations is designed so that the ultrasonic beam is directed from the side of a smooth shell at the end of a long flexible catheter. The transducer is introduced through a convenient vein into the organ to be examined. The display system is arranged so that the time-base starts at the centre of the cathode ray tube, and moves around radially as the transducer is rotated within the organ. Of course, the chances of detecting a smooth and extensive interface are very small, but it is claimed that small nodules and fibrous tissues can be easily displayed, because they reflect significantly through quite wide angles. Alternatively, this type of transducer may be used with a technique known as C-scan, where the probe is slowly withdrawn or inserted as the horizontal, brightness modulated time-base is moved vertically in sympathy. It is reported that intravenous transducers can give useful information in the liver and the heart.

(v) *Limitations*

The resolution requirements of tomographic systems, with the exception of the intravenous method which is a special case, are more complex than those of the A-scan technique. This is because the system must reproduce the echo from a particular point in the patient as nearly as possible at the same point on the display, irrespective of the direction of the ultrasonic beam. Two factors reduce the performance of a system in this respect. Firstly, the ultrasonic beam travels through different lengths of various media according to its direction, and velocity differences affect both the transit time and the direction of the pulse. Secondly, the stability of the transducer and patient supports control the accuracy. The best resolution that can be used is governed by the spot size of the display. In good quality displays, the spot diameter is about 0·1 cm. Because the display size is usually smaller than the anatomy which it represents, the acceptable positional performance of a system is commonly of the order of 0·25 cm.

Risk of Ultrasonic Damage

The mechanism of the various biological effects of ultrasound is the subject of considerable speculation. This is partly because the physical conditions of the various experiments which have been reported in the literature are not often clearly stated. In order to define the conditions in our work, we have devised measuring systems which are satisfactory at frequencies between about a half and ten megacycles per second, and powers between about half a milliwatt and fifty watts. The methods are described in papers published in 1963 and 1964.[1] The ultrasonic energy per pulse from a typical diagnostic instrument is of the order of two microjoules. From this figure, it is possible to compute the mean power at a given pulse repetition rate, and the peak power for a given pulse shape.

Ultrasound is used in medicine for three distinct purposes; these are therapeutic destruction, physiotherapy and diagnosis. It is used routinely to destroy the vestibular end organ in cases of Ménière's disease, and focusing transducers giving intensities measured in kilowatts per square centimetre have produced lesions deep in the exposed brain with applications of only a few seconds.

Like all forms of energy, ultrasound eventually turns into heat. Thus, any biological effect which can be produced by heat can also be produced, under the right conditions, by ultrasound. For this reason, it has been suggested that all ultrasonic effects on biological material are of thermal origin.

We have investigated the effect of ultrasound on the giant axon of the Squid. Plate II, Fig. 6 shows some of the results of a typical experiment. The dissected nerve was stimulated by means of an isolated electrical pulse, and the resultant action potential picked up in turn by two recording electrodes, connected through high input impedance amplifiers to the y-plates of a double-beam oscilloscope. Between the recording electrodes, the nerve dipped into sea-water, and an ultrasonic transducer was arranged to direct its beam through the nerve and into an efficient absorbing system. The temperature of the sea-water was monitored by a thermistor close to the irradiated section of the nerve.

In these oscillograms the upper trace is that due to the recording electrode close to the stimulating electrodes, and the lower trace is that due to the electrode at the other end of the nerve, beyond the region of ultrasonic irradiation. The sea-water was connected to the earth of the recording system. In Fig. 6(*a*), the stimulus level was just below the threshold required to produce an action potential, and the stimulus artifact can be seen in the upper trace. In the remaining figures, the stimulus

---

[1] The names of the authors appear in the list of References, and perhaps it may not be necessary to print them again here.

PLATE I

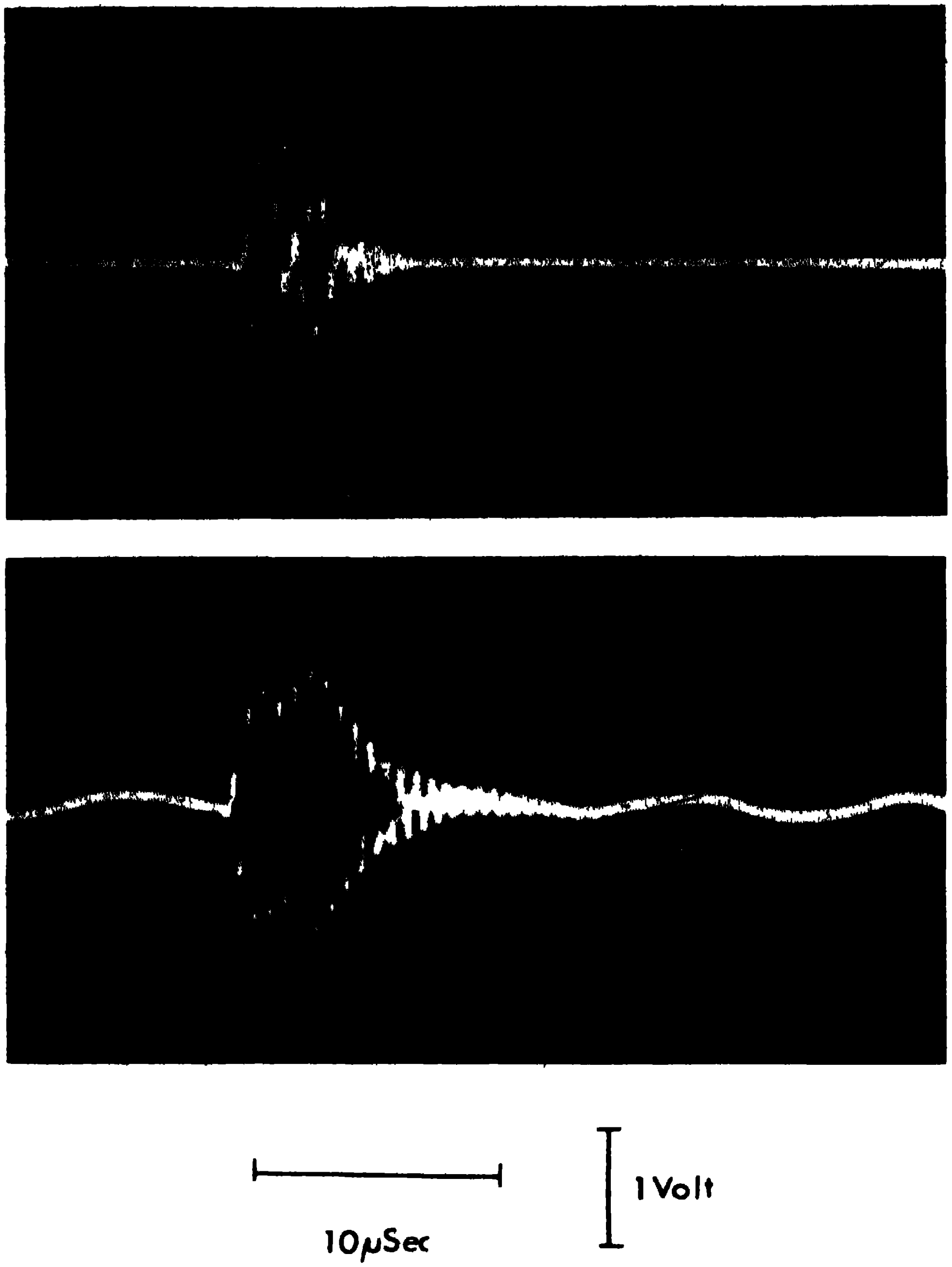


FIG. 3. Effect of transducer bandwidth on resolution. Echoes obtained from a perspex sheet 0.32 cm thick, at a range of 20 cm, using transducers of 2.0 cm diameter. *Upper trace:* damped transducer, 1.7 Mc/s; *lower trace:* undamped transducer, 1.8 Mc/s.

Plate II

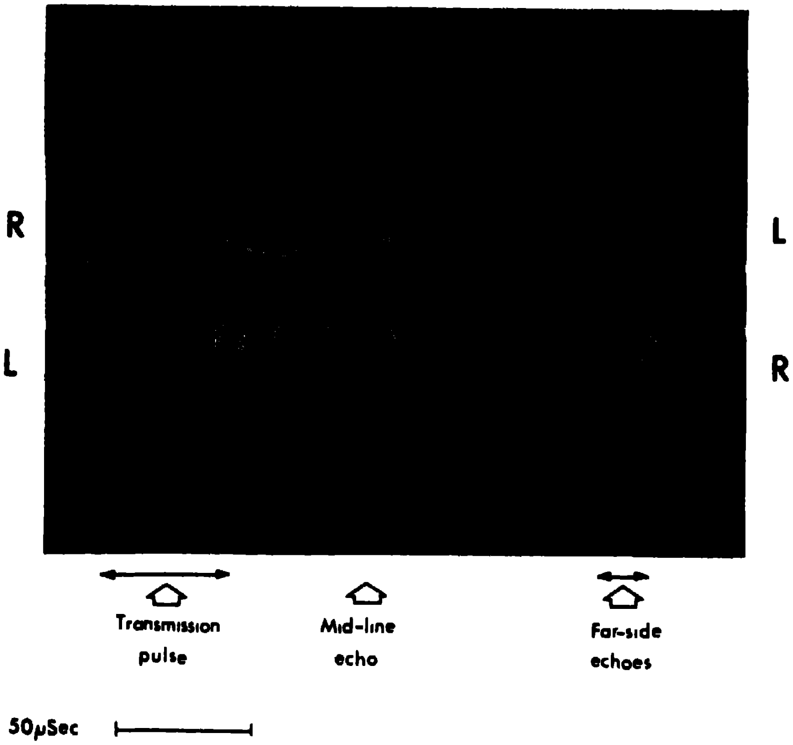


Fig. 5. Normal cranial echogram. Obtained using a 1.7 Mc/s, 2.0 cm diameter transducer.

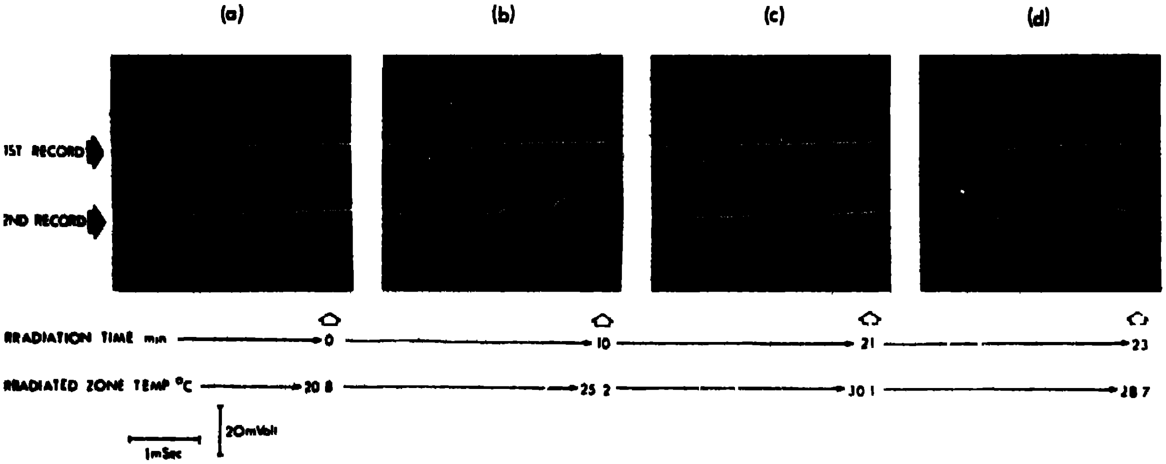


Fig. 6. Reversible nerve block induced by ultrasound. Oscillograms obtained with giant axon of Squid, *Loligo forbesi*, irradiated with ultrasound at 3 Mc/s, 25Wcm^{-2}. Distance between recording electrodes, 4.0 cm; length irradiated, 0.5 cm.

level was twice the threshold level. At the commencement of the experiment, the bath temperature was 20·8° C. After irradiation for 10 minutes the nerve was still conducting, although the temperature of the irradiated zone had risen to 25·2° C. During the 21st minute the temperature rose to 30·1° C., and conduction past the irradiated zone suddenly almost ceased. The ultrasound was left on, and by stirring the sea-water, the temperature fell to 28·7° C. The nerve started to conduct again. Because the irradiation was continuous, we concluded that the nerve block was due to the temperature rise caused by the ultrasound, and not to the direct effect of the ultrasound.

In another experiment, we measured the change caused by ultrasonic irradiation in the rate of sodium efflux from a giant axon. There is not space to describe this experiment in detail, but it seems probable that the increase in sodium efflux during irradiation was not incompatible with the effect of the temperature rise on the sodium pump.

These experiments show that it is often difficult to separate thermal from ultrasonic effects. However, it seems likely that cell damage associated with cavitation due to pressure changes in the ultrasonic wave is primarily the result of a mechanical effect. However, the intensity required to produce cavitation increases with the frequency, so that in water about 300 Wcm^{-2} are required at a frequency of 1 Mc/s.

Connolly (1963) has shown that cell damage is increased by the presence of bubbles introduced into a suspension of cells irradiated by ultrasound at an intensity much below that required to produce cavitation spontaneously.

Another possible mechanism for cell destruction is that the cell wall may resonate at the frequency of the ultrasound, thus accelerating the break-down of the membrane. Ackerman (1960) carried out experiments on various cell suspensions, and correlated the maximum breakdown rate with particular frequencies, according to the type of cell. However, he did not define the ultrasonic conditions, and his theory seems to have little support at the present time. Streaming of the cell contents has also been suggested as responsible for cell damage, and Nyborg and Dyer (1959) have reported the results of experiments using low frequency needle probes on the contents of single cells, but these experiments have not been extended and related to conditions at high frequencies in large masses of tissue.

A few workers have used medical echoscopes to carry out specific experiments on biological materials. Thus, Andrew (1964) reported no visible effect on the development of the embryo in frog and perch spawn irradiated for 24 hours by a commercial echoscope. Tsutsumi, Sano, Kuwabara, Takakura, Hayakawa, Suzuki and Katanuma (1964) irradiated the brains of dogs for various times with their echoscope. They measured the effect on two enzymes in the serum and C.S.F., and showed that there was evidence of reversible damage after several hours. However, the control conditions are not clearly defined, particularly with regard to anaesthesia.

My own view is that cell damage may be caused both by heat and by vibration, and that the relative importance of these factors depends upon the particular circumstances. At the frequencies employed in diagnosis, the most likely source of damage is heat, unless gas bubbles are artificially introduced; and the mean intensities used in diagnostic procedures cannot cause a significant rise in temperature. This view is supported by the fact that no damage attributable to ultrasound has been reported after several years of clinical use in diagnosis, and that the incidence of damage is quite low in physiotherapy, where the intensities are several thousand times greater, although the frequencies are much the same.

ACKNOWLEDGEMENTS

I am indebted to my colleagues, Mr. M. A. Bullen, Mr. H. F. Freundlich and Mr. J. Angell James, for much helpful advice. I wish to thank Dr. P. C. Caldwell of the University of Bristol, who collaborated with me in the experiments on the Squid giant axons: this work was carried out in the Laboratory of the Marine Biological Association, Plymouth. I also acknowledge the financial support of the Medical Research Council and the Board of Governors of the United Bristol Hospitals.

REFERENCES

ACKERMAN, E. (1960) *Proc. 3rd Int. Conf. Med. Electron.*, paper no. 25.

ANDREW, D. S. (1964) *Brit. J. Rad.*, **37**, no. 435, 185.

CHRISTIE, D. G. (1962) *Materials Research*, **86**.

CONNOLLY, C. C. (1963) M.Sc. Dissertation, University of Vermont, U.S.A.

DONALD, I. and BROWN, T. G. (1961) *Brit. J. Rad.*, **34**, no. 405, 539.

GORDON, D. (1963) *Med. Electron. biol. Eng.*, **1**, 51.

GREATOREX, C. A. and IRELAND, H. J. D. (1964) *Brit. J. Rad.*, **37**, no. 435, 179.

HOWRY, D. H. (1955) In *Ultrasound in Biology and Medicine*, ed. E. Kelly, p. 49. Washington: Institute of Biological Sciences.

KIMOTO, S., OMOTO, R., TSUNEMOTO, M., MUROI, T., ATSUMI, K. and UCHIDA, R. (1964) *Ultrasonics*, **2**, 82.

KOSSOFF, G., ROBINSON, D. E., LIU, C. N. and GARRETT, W. J. (1964) *Ultrasonics*, **2**, 29.

LANGEVIN, P. (1920) *Brit. Pat. Specifications*, N.S., **457**, no. 145691.

LEKSELL, L. (1955) *Acta chir. scand.*, **110**, 301.

NEWELL, J. A. (1963) *Phys. Med. Biol.*, **8**, no. 3, 241.

NYBORG, W. L. and DYER, H. J. (1959) *Proc. 2nd Int. Conf. Med. Electron.*, p. 391.

TSUTSUMI, Y., SANO, K., KUWABARA, T., TAKAKURA, K., HAYAKAWA, I., SUZUKI, T. and KATANUMA, M. (1964) *Med. Electron. biol. Eng.*, **2**, 21.

WELLS, P. N. T., BULLEN, M. A., FOLLETT, D. H., FREUNDLICH, H. F. and JAMES, J. A. (1963) *Ultrasonics*, **1**, 106.

WELLS, P. N. T., BULLEN, M. A. and FREUNDLICH, H. F. (1964) *Ultrasonics*, **2**, 124.

WILLOCKS, J. (1962) *Proc. Roy. Soc. Med.*, **55**, 640.

IV

Sophisticated Laboratory Animals

W. LANE-PETTER

THE oldest meaning of the word sophisticated given in the Shorter Oxford English Dictionary is 'altered from, deprived of, primitive simplicity or naturalness (1603)'. Later definitions introduce a pejorative note, suggesting adulteration, corruption, perversion or falsification. The animals that I am going to discuss have certainly been deprived of their primitive naturalness: deprived, that is, as far as practicable of genetic imprecision, disease, infestation, nutritional deficiency and other stressful influences. They are animals whose sophistication implies removal of adulteration, corruption, perversion and falsification: and thus the word sophistication in its present sense has undergone a reversal of meaning.

Twenty years ago it was known what sort of animals were needed. 'Pure lines', pathogen free and germ free animals, and special strains were all known to exist, and their potential usefulness could be deduced from actual experience. What had to be discovered was how to provide such animals in the requisite numbers, practically and economically. And not only to provide them, but to distribute them efficiently, avoiding wasteful overproduction and periods of famine. There had to be an investigation of factors governing the demand for animals of various kinds, so that useful predictions of needs could be made. These were not problems of exceptional difficulty, but they did call for a scientific approach, for in no other way could a solution be expected.

However, before discussing the logistic problems, I would like to show the directions that sophistication has taken.

The Demand for 'Pure Lines'

The term 'pure line' has been used in the past as if it were synonymous with inbred strain. 'Inbred strain' was itself ill defined, being used to describe anything from a colony with strict brother × sister mating to nothing more exacting than a closed population. The first formal definition appeared in 1952 (Committee on Standardized Nomenclature) and was as follows: 'A strain shall be regarded as inbred when it has been mated brother by sister (hereafter called b × s) for twenty or more consecutive generations. Parent × offspring matings may be substituted for b × s matings, provided that in the case of consecutive parent × offspring matings the mating in each case is to the younger of the two parents.'

Professor Grüneberg once told me that he accepted this definition (indeed, he was one of the authors of it) but would like to stipulate that the breeding be managed by someone in whom he had confidence, and this, the mark of authenticity, is an essential part of any definition, and one that may be overlooked. The authentication of descriptions and designations is something that arises over and over again in this field.

An inbred strain, so defined, possesses maximum homozygosity. The animals of such a strain are within all practical limits genetically identical; tissues will transplant successfully from one animal to another, and gene-dependent characters will appear with predictable regularity.

Medawar (1958) has pointed out that for twenty years or so cancer research was seriously handicapped because no genetically pure strains of animal were available for experiment. The absence of the most appropriate experimental material, namely the inbred mouse, had led, in his view, to three false trails being pursued. Without such strains it was impossible to differentiate between acquired and natural immunity to tumours. Secondly, it was impossible to distinguish tumour immunity from tissue immunity. Thirdly, inbred strains were essential in demonstrating that the fate of a tumour, spontaneous or transplanted, was partially or totally dependent on the genetic make-up of the host animal.

But strains of maximum homozygosity have many disadvantages. They lack vigour in comparison with non-inbred strains, and this expresses itself in such characteristics as lower productivity, poorer growth, and a greater lability of response to minor defects or fluctuations in the environment. Their genetic authentication must also be carefully established, and this is a matter of some technical skill, for accurate records have to be kept and a single wrong mating may ruin the work of years. Subline formation, through the excessive persistence of parallel lines; divergence, which is the beginning of the same process; genetic drift; and the accidental loss of useful characteristics (such as a high incidence of leukaemia) through mutation, are all likely to happen in a colony that is not expertly managed.

Not to be overlooked also are the first generation crosses between two inbred strains. These seem to have the best of both worlds, for they are all as genetically identical as the parent colonies from which they derive, but being crosses they have a hybrid vigour that their parents lack. For many purposes they are the most suitable animal, although they do not fulfil all the claims that have been made for them.

Histocompatibility, special characteristics such as a high incidence of mammary tumours, and enhanced lability of response are the main features of inbred strains that make them useful, even essential. This lability of response, a sort of unstable equilibrium which is characteristic of the homozygous state generally, is for the most part non-specific. For many years we maintained our inbred strains of mice in wooden boxes (of a type that could be repeatedly sterilized) because we had found that in the alternative available, namely metal boxes, their breeding performance was significantly depressed. The same was not true of non-inbred mice. The advent of plastic boxes has enabled wood to be abandoned, plastics approaching wood in being quiet to handle and having a low conductivity of heat, being very much better than metal in this respect. More recently we have exploited lability of response in assessing diets; the CBA strain of mouse is a more sensitive indicator of dietary adequacy than any other strain we have investigated (Porter, Lane-Petter and Horne, 1963).

However, gene-dependent characteristics are not only to be found in inbred strains and their crosses. Everyone has heard of Sprague-Dawley, Wistar, Long-Evans, Glaxo, and Charles River rats; of Swiss, CFW, Schofield, LAC Grey mice, and many others, and most of them are not inbred. Investigators rightly insist on their own choice of strain and designate it in their published reports. The weakness, however, lies in the absence of exact definitions in most cases, and often in the doubtful authentication of sources. Thus, the terms Swiss and Wistar have come to mean little more than albino mice and rats respectively.

Nevertheless, there are important and consistent differences in response that characterize most of the named non-inbred strains. These are exactly comparable with the differences between breeds of dog or other domestic animals, and they arise from the same cause, namely the presence of homozygosity in the strain in respect of a certain number of genes that control characteristics of the strain, while stopping short of the highly homozygous condition of the inbred strain.

Characteristics may thus be fixed by limited homozygosity in an otherwise heterozygous strain. The anaphylactoid reaction to injection of dextran in rats is probably dependent on a single gene (Harris, Kalmus and West, 1963) and very many pathological mutants in the mouse are monogenic in origin (see Grüneberg, 1956). Such genes may be carried in homozygous state in otherwise heterozygous animals, so avoiding many of the practical disadvantages of inbred strains.

The dependence of specific responses on one or perhaps a few genes is quite general. Dwarf beef cattle, which are homozygous recessives for the dwarf gene, were found to be more sensitive to insulin than homozygous normals or heterozygotes (Foley, Heidenreich and Lasley, 1960). Thalidomide had no effect on pregnant C57BL mice, but led to considerable resorption of foetuses in the A strain (Hagen, 1963). Similarly, trypan blue had significantly different effects on foetal resorption and damage in pregnant rats of three different strains, two of them non-inbred (Gunberg, 1958). In twelve strains of mice given a standard dose of hexobarbital, there was an almost threefold difference in sleeping times between the most and the least sensitive strains (Jay, 1955). Pseudopregnant rabbits of the New

Zealand White strain (a non-inbred strain) lactated when given reserpine intravenously, but another non-inbred strain, the Dutch, did not lactate in response to the same treatment (Tindall, 1960). McClearn (1962) has found that some strains of mouse are excited by alcohol, some are depressed, some are little affected and some absolutely refuse to take it (Rogers and McClearn, 1962).

Perhaps the most dramatic of all examples is that of my colleague, Dr. Annie Brown (1962), who found that sensitivity of mice to histamine acid phosphate, as measured by the LD50 in sensitized animals, varied from the most sensitive to the least sensitive strain by a factor of 50 or more.

The genetic possibilities of our familiar laboratory species have hardly begun to be exploited, and an almost unsuspected differentiation in the pharmacological field is only now coming to light. The new discipline of pharmacogenetics has been reviewed by Kalow (1962), Meier (1963), Brown (1963, 1964) and others.

THE PROBLEM OF DISEASE

Ordinary laboratory animals carry a collection of infections, some of which give rise periodically or continually to overt disease, while others remain silent except perhaps when they are specifically provoked. They also carry helminths and ectoparasites. Clearly, it is desirable to rid the animals of all possible causes of disease by removing from them their pathogens or potential pathogens.

Attempts have been made in the past to eliminate infections in various ways: by selection of non-infected individuals and breeding them in isolation, by chemotherapy, by vaccination, by raising the quality of the environment, and by many other means. Such methods have been frequently successful with certain infections, notably those of bacterial causation, but they have little or no effect on most virus infections, nor on some parasitic infestations, such as pinworms (*Aspicularis*, *Syphacia*) in mice. Total elimination from all animals at one time is necessary if the infection is to be excluded, coupled with precautions against reinfection.

Chemotherapy—even against endo- and ecto-parasites—is seldom totally effective, and a single failure may be sufficient to re-establish the infection. Selection can be more successful, but it is a tedious and prolonged process, and it depends for its success on two conditions, namely a reliable diagnostic test of infection, and an incidence of infection in the colony of less than 100 per cent. Not every infection is obliging enough to fulfil these conditions. Vaccination of laboratory animals is probably the last refuge of the desperate. It may be useful in controlling savage epidemics, but its value in a programme of elimination is doubtful in all infections for which it has been tried. Improved husbandry and hygiene can probably contribute as much as any of these methods to cutting down the incidence of disease in a laboratory animal colony, and in a well-run colony they should certainly be of a high enough order to prevent the introduction of infection from outside. But with all these methods, success may only be partial. Infantile diarrhoea, ectromelia, pinworms and mites of mice; chronic respiratory disease of rats; coccidiosis and respiratory infection of rabbits; possibly salmonellosis of guinea-pigs; are among the diseases that are likely to remain when all such attempts have been made to eliminate them.

Fortunately, new techniques borrowed from the students of germ free life have altered the whole picture. A full-term foetus removed aseptically from the mother and raised in a strictly aseptic environment will possess no micro-organismal contamination apart from what may have crossed the placental barrier *in utero*. In mice, rats, guinea-pigs and rabbits nothing bigger than a virus, and only some of these, can cross this barrier, so that this becomes a method of eliminating, at one stroke, virtually all resident infections, pathogenic or not. But such monobiotic, germ free, axenic—call them what you will—animals only stay germ free in an aseptic environment, such as an isolator. Expose them to the air of even the cleanest laboratory, and they acquire a mixture of bacteria and viruses to which they may or may not succumb. If they survive—and rats and mice usually do—they rapidly establish a balance with their microflora, and if this flora includes no pathogens, then they continue to be free from disease—in other words, just what the investigator wants them to be—and they remain so unless or until they become pathogenically infected from another source. But such animals, though free from disease, are radically different from what they were before being removed from the sheltered life of the isolator, and the change they have undergone is irreversible.

Raising mice and rats by calling into use these refined techniques is becoming common practice today. It is technically feasible, not too difficult and, surprisingly, not uneconomical. So much so, that none of us today can afford to raise our rats and mice, and in the near future other animals as well, in any other way, because the investigator, once he learns the advantages of these so-called 'SPF' animals, will rightly demand them all the time. 'SPF' stands for specific pathogen free, and is an impossible term to define, because it is not an entity but denotes animals in that state of health that we would like them always to enjoy. Indeed 'SPF' is merely a convenient term of passing value, implying three conditions: aseptic derivation by hysterectomy, with hand rearing or fostering on to 'SPF' mothers: maintenance within a barrier that is, as far as practicable, proof against the accidental invasion of the colony by pathogens: and skilled care, without which the animals will not thrive and the barrier is certain to break down. These are advances in technology that have enabled us to achieve quickly what at one time seemed so far beyond our reach. There is a tendency for the dramatic success of derivation by hysterectomy and barrier maintenance in cleaning up animal colonies to overshadow this third and equally important requirement of skilled care. But this tendency must be resisted. You do not put untrained, still less untrainable, labour into your instrument workshop. Equally, it would be self-defeating to employ such labour in a sophisticated animal house. Such a place must be manned by trained animal technicians.

'SPF' animals are, in fact, conventional animals less their common and troublesome diseases (Lane-Petter, 1962); in other words, the healthy animals that have been demanded and sought for so many years. But they remain conventional animals in all senses that interest the investigator, and they do not die of irrelevant conditions when stressed.

The question therefore does not arise, are 'SPF' animals going to be extensively used in the future? For scientific work an animal that does not have an unknown burden of pathogenic infection must be preferred to one that does; just as the chemist prefers clean glassware to dirty. The cleaning of animal stocks is a technical exercise, the difficulty and expense of which have been much exaggerated—perhaps because novelty is always remarkable and usually expensive. Experience has shown that reinfection does not readily take place, even in a good conventional animal house. The breakdown of barriers, once established, has been a rare occurrence.

It is important to be quite clear about this, because so much confusion has arisen regarding the nature and relationships of germ free, gnotobiotic and 'SPF' animals. Germ free animals have no microflora, apart from what viruses have crossed the placental barrier. Gnotobiotic animals have a microflora that is known in its entirety, because it has been deliberately implanted in conditions absolutely controlled in the bacteriological sense, usually by inoculation or contamination of germ free animals. (The term gnotobiotic is also used to include germ free, a use which would seem to be quite correct.) Both germ free and gnotobiotic animals exist in isolators and only in isolators. It is possible to define with certainty the forms of life—host and microflora if any—existing inside an isolator, because the isolator walls are an absolute barrier against micro-organisms. But 'SPF' animals are not inside such a barrier. They or their ancestors have been separated at Caesarian birth from those infections of the mother that have not passed through the placenta but which she would be likely to donate to them in the nest, were she to rear them; and they are raised and subsequently bred in an environment that is not absolutely germ proof, like an isolator, but where the entry and survival of pathogens is unlikely and discouraged to the point of despair. Such animals pass through the neo-natal period in relative safety, and do not acquire the infections conventional nurslings almost invariably acquire. When, at or after weaning, they are taken into the less protected world of the experimental animal house, they prove at least as capable of coping with infective hazards as their less healthy conventional counterparts.

In the Report of the Commission on Drug Safety (1964) it is recorded that, for basic research on toxicity 'strong opinions against the use of pathogen free rats were voiced; these rats have severe liabilities with respect to incidence of spontaneous tumours and lack of resistance to infection. . . . Normal, hardy animals were preferred'.

The only explanation for the extraordinary reference to severe liabilities is that the Committee was confusing pathogen free with germ free, because the statement is just not true. But then, what are 'normal, hardy animals'? Are 'normal' rats those that regularly develop chronic and severe lesions of the respiratory tract at two years of age; that carry bartonella, salmonella, ectoparasites, helminths, middle ear infection; and are they hardier than their healthy counterparts? Are 'normal' mice those from which the virus of ectromelia can regularly be isolated by sufficiently persistent effort and which 'normally' have mites, worms and an unknown complement of mouse pathogens? Does the investigator really prefer to work with animals that are a ragbag of endemic and epidemic infection, with uncertain viability and non-reproducible reactions? Are his results more reliable because he uses contaminated reagents, dirty glassware and wrongly calibrated instruments? If 'normal' has any meaning in this context, it must mean unstressed; that is, totally free from all stresses such as may arise from infection, malnutrition, poor physical environment, bad care and handling and so on: because any stress is bound to affect the animal, and an unknown type and amount of stress will affect it in an uncertain way. At the same time, it can be argued that an unstressed animal is also necessarily abnormal, so probably the nearest approach to the concept of a normal animal is that highly abnormal creature, the germ free animal, which is the very antithesis of what is contemplated by those who talk glibly about 'normal' animals. A plea has been made for healthy animals; these are now available, but there is still, in a few quarters, a reluctance to use them. They are marginally more expensive to produce, but there is virtually no wastage from intercurrent disease; and the saving in effort because experiments do not break down is incalculable.

NEW OPPORTUNITIES

The elimination of murine epidemic disease from laboratory animal colonies may be compared with the elimination of plague, typhus, typhoid, diphtheria, tuberculosis and many other diseases from human communities. Moreover, it has made possible the exploitation of strain differences, many of which were previously masked or nullified by variability due to disease or poor nutrition and care. Indeed, this has underlined the point made twenty or more years ago, that the choice of animal, of strain as well as of species, can be crucial. Williams (1964) has drawn attention to species differences, and shown that the analogies between animals and man seem to follow no pattern based on close zoological relationships. Within a single species, the mouse for example, there are also differences, and these can be of the same order as those existing between species. Thus, there can never be an answer to the question, what is the best animal, species or strain for use? It depends on what it is to be used for.

At a recent meeting in Hanover there was an entertaining excursion into laboratory animal mythology. Various speakers referred to the 'universalversuchstier', the 'allzweckmaus', the 'all purpose mouse' and the 'polyvalent hybrid'—blood relations, no doubt, of Goodwin's 'ideal experimental animal' (1963). These animals have no more reality than the gryphon, the roc or the leprechaun. To go chasing them may be a lot of fun, but it is to ignore the very real opportunities that are before us of exploiting the sort of differences Williams and the pharmacogeneticists describe.

To do so, it is necessary to understand clearly what an experimental animal is. It is, in the main, an analogue, that is, a system intended to simulate another system (man, for example) that cannot be observed or manipulated directly. Because our animals are breeding so much better, and are more productive of useful results than ever before, we are tempted to use them ever more extensively. Unless we are more self-critical in our use of animals, in our choice of species and strain and quality and in our experimental design, there is a danger that the resources of the animal department will outstrip the resources of the laboratory where the animals are to be used.

Before an animal is used for an experiment, three questions must be asked. First, is the animal likely to be the best experimental system to provide an interesting and meaningful answer? To use mice to test for the presence of lead in drinking water might provide an answer, but it could be more easily obtained by a simple chemical test. The second question is, what sort of animal should be used? No animal is a perfect analogue: the choice may depend on size (dogs in experimental surgery), short life span (rats for tests of carcinogenicity), specific susceptibilities (guinea-pigs for detecting tubercle bacilli): and today the choice is so wide that the investigator will often need expert help in seeking the best subject for his work. The third question is, how many animals are needed? From the large number demanded by the statistician, who has shown how to rescue useful information from experiments that went wrong because some of the animals died, we come to the single animal advocated by the same group who repudiate 'pathogen free rats'. They say 'It is more important to study individual animals in depth than to study large numbers superficially' (Report, 1964). But surely, by the individual animal it is not meant to imply that one animal is the beginning and end of the whole investigation. And if the animals from which this specimen is chosen are afflicted with a variable assortment of lesions, infections, deficiencies, developmental peculiarities and genetical uncertainties, and if the individual specimen that happens to come out of the bag is too afflicted to survive investigation in depth, are we any the wiser? To study animals in depth only becomes possible when there is some constancy, some uniformity, about the animals being studied.

The circumstances of the actual experiment are also important, sometimes crucially so. As Russell and Burch pointed out in their penetrating examination (1959) the experimental animal is the product of its genotype, its environment and what D'Arcy (1962) refers to as its conditioning, both immediately before and in the course of the experiment. Thus the genotype and the environment interact to produce the phenotype, and the phenotype is subjected to conditioning to create the dramatype. It is the dramatype that constitutes the experimental animal. The importance of this concept lies in recognizing the fact that animals are not, like chemical reagents, there on the shelf, unchanging until the time comes for them to be used. They are always changing, always reacting to the world around them. However perfect the environment of the breeding or stock rooms of the animal house, the time comes for the animals to be prepared for experiment. This means more handling (not necessarily a disturbing stimulus), perhaps transfer to a new type of cage; a new room, a different technician in charge, probably rearrangement of social groupings; different noises, smells and even diets. Fortunately, mice and rats are among the most adaptable of all mammals, other than man, and that is why they are such outstandingly successful laboratory animals; but there are limits to the stress of altered conditions that even mice and rats can tolerate with equanimity, and these limits may be overstepped.

The Choice of Animal

The choice of animal before the investigator today is dazzling. There are germ free animals, raised in isolators, and dissociated from all other forms of life at least down to medium sized viruses. There are gnotobiotic animals, which also have to live out their lives in isolators; they are contaminated with known micro-organisms. Because both germ free and gnotobiotic animals maintain their respective conditions only so long as they remain in an efficient isolator, any experiment they undergo must be carried out within the isolator. This makes them expensive and inconvenient to use as well as to produce, but for certain purposes they have and will continue to have an essential use. They may be compared with the early development of tissue cultures, which used to be regarded as so difficult and troublesome that they should be avoided at almost all costs. But improvements in technique made of tissue culture a new and vitally useful tool, which has found its way into many biological disciplines. Today germ free isolators are becoming cheaper, simpler and incomparably easier to operate, and before long it will be possible to buy a disposable isolator complete with a cage of animals and enough sterile food, water and bedding inside it to last for up to a month. Such outfits will have a limited but necessary place in research, like tissue culture, but none would willingly use a germ free outfit, or a tissue culture, if open—that is, non-germ free—animals or agar plates would do.

Then there are 'SPF' animals, the circumstances of whose production makes it exceedingly unlikely—but not impossible—that they are associated with any epidemic or endemic pathogen. They may, and usually do, carry potential pathogens, such as *Escherichia coli*, *Staphylococcus aureus* and others, but they may be expected not to die of intercurrent disease, and in practice it is found that they very seldom do. And of course there are the animals we have all suffered from—mice with ectromelia, salmonellosis, Tyzzer's disease, infantile diarrhoea, ectoparasites and endoparasites; rats with chronic respiratory disease and bronchiectasis, leptospirosis, nephritis, scabies and ringtail; guinea-pigs with scurvy, pasteurellosis, salmonellosis and cervical adenitis; rabbits with coccidiosis, mucoid enteritis, ear canker and snuffles—and so the disreputable story could go on.

Continuing to summarize the choice available, there are inbred strains, and F1 crosses between inbred strains, which are genetically uniform but have the advantage of hybrid vigour. There are non-inbred strains, usually possessing certain genetically fixed characteristics, and also those carrying specific mutants.

All this refers in the main to mice and rats, but of course these are not the only laboratory animals, although they account numerically for some 85–90 per cent of all vertebrates used. Hamsters and other rodents, guinea-pigs and rabbits, dogs, cats, pigs and monkeys, and from time to time less usual laboratory species, are all available. Hamsters, for example, already offer inbred strains and guinea pigs pathogen free versions, and there is no reason why other species should not be developed in this way: indeed, many of them are. But the mouse and rat maintain their lead, and must continue to do so, because we know so much about them. Other species will only be used for purposes which the mouse and rat cannot be made to serve.

Provision of Sophisticated Animals

When a mouse was just a mouse, a live specimen of *Mus musculus*, or with spurious precision a 'white mouse', anybody who could make the species reproduce itself could expect to sell the progeny to a laboratory. And since it is not difficult to make mice and rats and guinea-pigs and rabbits and hamsters breed, almost anyone could do it. And almost anybody did, either commercially as backyard breeders, or in the laboratory animal house as an ill-favoured subsidiary occupation. But this sort of casual inexpert breeding cannot produce the more sophisticated animals that are needed today.

Whether the animals to be produced are genetically defined, or of defined health status, a key requirement is authentication. As the contents of the bottle must correspond exactly with what is on the label, so the defined animal must be what it is said to be. An inbred strain must be produced in such a way as to merit Professor Grüneberg's confidence. 'SPF' animals are no better than any others if the conditions of their rearing do not make the accidental acquisition of pathogenic infection a vanishingly rare event. Control measures, such as routine skin grafting of inbred strains and a constant search for trouble in disease-free colonies, must be the rule, and they must be efficiently and expertly conducted. And this raises a fundamental logistic problem.

Experience and skill in the application of these control measures are specialized accomplishments, and they will only be found where ample opportunities exist for their regular exercise. Such skills tend to be concentrated in centres devoted to the study of laboratory animals in their own right, although I hasten to add that this is by no means intended to be an exclusive claim. It is a tendency, enough to indicate that a laboratory, just by virtue of being a laboratory, is not *ipso facto* competent to apply the control or exercise the skill needed to produce sophisticated animals. The chemist does not have to be expert in chemical manufacture, nor the engineer in metallurgy: apart from their own subjects all they need to know as a rule is their own limitations. So with the experimental biologist.

The pattern that is growing up is to concentrate production of sophisticated animals in places that are big enough to practise advanced techniques of large-scale breeding and efficient control services. But since concentration of production carries great hazards, both genetical and infective, a breakdown into two stages must be arranged. I have discussed this in some detail elsewhere (Lane-Petter, 1961) and do not wish to do more than reiterate the system as a general principle which appears to be of wide if not universal application in laboratory animal production. The first stage is the *primary type colony*, where control is maximal, and whose function is to provide breeding stock for limited subcultivation in the second stage, namely one or more *production units*, where animals for experiment are bred.

The principle of two stage production, emerging from antecedent chaos, has made possible the provision of sophisticated laboratory animals wherever, and in whatever numbers, they are required. This in turn is throwing light on the contribution that experimental animals have to make to the further advance of knowledge. If I may, with abject apologies to Gertrude Stein, express it in a loose and literary way, a mouse is not a mouse is not a mouse is not a rat is not a dog is not a monkey is not a man. Worse still, a CBA mouse is not a C3H mouse, although they look alike. But sophisticated animals are more constant in at least some of their responses, and results obtained with them more regularly reproducible on others of their kind. It is thus possible to explore them in greater depth, to examine the reaction of the whole animal as profoundly as the physiologist examines the reaction of an isolated tissue or organ. I would expect to see a decline in the use of the relatively crude LD50 test for toxicity in favour of assessment of more subtle effects than death, on selected types of animal known to have, for example, specially appropriate enzyme systems. This is the same sort of difference as exists between obliteration and target bombing: polypharmacy and specific therapy: custard pie slapstick and high comedy.

There are those who deplore the use of animals for experimental purposes, and they include even some scientists whose work compels them to it. Others, including myself, are grateful that legitimate intellectual pursuits compel us to make use of the opportunities that the experimental animal presents, and only deplore the fact that sometimes this use causes the animal to suffer, and thus perhaps turns a use into an abuse. But our lives as biologists are enriched because the study of animals, for its own sake, is forced upon us by our pursuit of knowledge. It is no more to be deplored that we use animals for experiment than that we raise animals for food, work or any other legitimate purpose. As for the animals, most of them would not even exist if they had not been bred for our purposes, and therefore they owe their very lives to the use we make of them. I submit that we have an obligation to use them to the best advantage. In doing so we not only follow our biological callings to the full, thereby reaping the richest possible harvest of new knowledge; but we also contribute, in a way that no animal protectionist can ever do, to the reduction, perhaps the eventual elimination, of suffering among the experimental animals to which we owe so much, and are on the way to owing so much more in the future.

REFERENCES

Brown, Annie M. (1962) *J. Pharm. Pharmacol.*, **15**, 406.

Brown, Annie M. (1963) *Animals for Research*, ed. W. Lane-Petter, Academic Press, London, p. 261.

Brown, Annie M. (1964) *Evaluation of Drug Activities: Pharmacometrics*, ed. D. R. Laurence and A. L. Bacharach, Academic Press. London, p. 111.

Committee on Standardized Nomenclature for Inbred Strains of Mice (1952) *Cancer Res.*, **12**, 602.

D'Arcy, P. F. (1962) *J. Pharm. Pharmacol.*, **15**, 411.

Foley, C. W., Heidenreich, C. J. and Lasley, J. F. (1960) *J. Hered.*, **51**, 278.

Goodwin, L. G. (1963) *L.A.C. Collected Papers*, **12**, 5.

Grüneberg, H. (1956) *Medical Research Council Memorandum, No. 33*, H.M.S.O., London.

Gunberg, D. L. (1958) *Anat. Rec.*, **130**, 310. (abs.)

Hagen, E. Odette (1963) *Lancet*, **1**, 501.

Harris, J. M., Kalmus, H. and West, G. B. (1963) *Genet. Res. Camb.*, **4**, 346.

ay Jr., G. E. (1955) *Proc. Soc. exper. Biol. Med.*, **90**, 378.

Kalow, W. (1962) *Pharmacogenetics*. W. B. Saunders, Philadelphia, U.S.A.

Lane-Petter, W. (1961) *Provision of Laboratory Animals for Research*. Elsevier, Amsterdam.

Lane-Petter, W. (1962) *Proc. Roy. Soc. Med.*, **55**, 253.

McClearn, G. E. (1962) *Proc. 3rd Int. Conf. on Alcohol and Road Traffic*, London: Brit. Med. Ass.

MEDAWAR, P. B. (1958) *The Harvey Lectures 1956–1957*. Academic Press Inc., New York.

MEIER, H. (1963) *Experimental Pharmacogenetics*. Academic Press, New York.

PORTER, G., LANE-PETTER, W. and HORNE, NANCY (1963) *Zschr. Versuchstierk*, **2**, 75.

REPORT OF THE COMMISSION ON DRUG SAFETY (1964). 29 and 31. Washington, D.C.

ROGERS, D. A. and MCCLEARN, G. E. (1962) *Quart. Jl. of Studies on Alcohol*, **23**, 26.

RUSSELL, W. M. S. and BURCH, R. L. (1959) *The principles of humane experimental technique*, Methuen, London, p. 115.

TINDAL, J. S. (1960) *J. Endocrinol.*, **20**, 78.

WILLIAMS, R. T. (1964) In *Proc. Europ. Soc. Study Drug Toxicity*, **4**. Excerpta Medica Foundation, Amsterdam.

V

Demography and the Cancers

R. A. M. CASE

DOLL (1965), speaking about the possibility of controlling various forms of cancer by preventive measures at a symposium held at the 1964 Annual Meeting of the British Medical Association, said:

> Cancer was once regarded as an inevitable accompaniment of ageing. Now the position has altered dramatically and cancer prevention has become a major function of public health.
>
> This change is due to four developments. First, we now know that the common cancers vary in incidence from place to place and time to time. The causes are only partly known, but experience of different countries shows that modern civilization is compatible with a relatively low incidence of nearly every type of cancer. Secondly, it is realized that cancer does not usually arise *de novo* from an otherwise normal tissue, but is preceded by the appearance of a series of histological changes proceeding through carcinoma-*in-situ* to invasive cancer. If detectable these changes provide an opportunity for effective treatment, or perhaps reversion of the process, by removal of the causative agent. Thirdly, knowledge of factors that affect individual susceptibility—genetic, constitutional, and psychological—is beginning to accumulate so that prophylactic efforts may possibly be concentrated where they will be most rewarding. Fourthly, there are the immense developments in virology.

And in the last sentence Doll might well have added immunology to virology.

It is my purpose to try to outline how demographic studies have contributed to these developments, and to suggest what extension of such studies might profitably be made in the future.

It is, of course, impossible in the space allotted to me to deal with all the major contributions to the subject, so I shall confine myself to some special aspects that have appeared to me to be seminal, or to some topics with which I and my colleagues have been closely concerned.

The fact that I shall not discuss other approaches to the cancer problem than the demographic one does not imply that I consider them to be of lesser importance. At all stages of our efforts to understand, control or treat the disease a constant interchange of ideas between the demographic, the clinical and the experimental disciplines is essential.

WHAT IS DEMOGRAPHY?

What do we mean by demography? For our present purpose, and, indeed, for most purposes of human biology, I do not think that we can do better than to refer to Sir Arthur McNalty's (1961) *British Medical Dictionary* which defines demography as 'The social science of people considered collectively: race, occupation, habitation, physical, moral and intellectual conditions, and vital statistics.' This definition is subdivided as follows: 'Static demography: The collective study of the anatomy of a community, e.g. Numbers, sex, age, social position, occupation, environment,' and 'Dynamic demography: The collective study of conditions within a particular community with reference to the vital statistics considered over the period of the duration of the community.'

DEVELOPMENT OF RECORD KEEPING

In general, medical demographic studies rely on the recording of two kinds of information, one kind referring to the people to whom an event may happen, commonly called the 'population at risk', and one kind referring to the event under consideration, in our case this being the development of or death from some form of cancer.

The history of these two kinds of recordings has taken somewhat different paths, and so will need to be considered separately.

The enumeration and description of populations is of considerable antiquity, early examples are to be found in the fourth book of Moses, the chronicles of the time of Augustus Caesar, and to some extent in our own Domesday Book.

Oddly enough, the *deaths* in the great plague of London were ultimately responsible for focusing attention upon the need for information about the *living* which nowadays is obtained from a periodic census. The plague year of 1592 saw the institution of the practice of issuing accounts of burials, or bills of mortality, for the City of London. The continuance of the practice of issuing these bills other than in plague years enabled John Graunt, who must be regarded as the father of modern demography, to publish his 'Natural and Political Observations. . . . Made upon the Bills of Mortality' (Graunt, 1662).

Graunt studied the records of twenty years between 1629 and 1658, omitting some years when plague was rife, in great detail. He remarks: 'in 20 years . . . there died 229,250'. In his itemized list of deaths he records 'Cancer, Gangrene and Fistula, 609 (deaths)'. Apart from indicating that cancer was then recognized as a cause of death worthy of mention, these figures are themselves of little concern to us, for they cannot possibly represent the true mortality from what we would by modern standards regard as cancerous diseases.

However, at the end of the fascinating deliberations which led him to devise, amongst other things, one of the earliest life tables, Graunt writes: 'It may now be asked, to what purpose tends all this laborious buzzling, and groping?' And he itemizes various objects of knowledge including the following:

1. The number of the People?
2. How many *Males*, and *Females*?
3. How many Married, and Single?
4. How many *Teeming* Women?
5. How many of every *Septenary*, or *Decad* of years in *age*?
9. What proportion die of each general and particular *Casualties*?
10. What years are Fruitfull, and Mortal, and in what Spaces, and Intervals, they follow each other?
13. Why the *Burials* in *London* exceed the Christenings, when the contrary is visible in the Country?

These items are remarkable for their prescience. They accurately foreshadow the need for the information which, after more than a century and a half, was to be demanded of the public, and which has served as the basis not only of population statistics and population projection but also of the classical pattern of the study of vital statistics in nearly all its present form. In them we have the concept of a population broken down by age and sex, and the basic requirements for fertility analysis, the need for tabulations of causes of death, the idea of a changing pattern of fertility and mortality, and a clear statement that there may be an urban/rural gradient in mortality.

In 1801 the first official census of Great Britain was taken, but it was not until 1821 that questions about age were asked, and then only on a voluntary basis. The question was not repeated until the census of 1841, but has remained in some form at each census since then. However, it is to be noted that even now a statement of the exact date of birth is not required.

We must now turn to the recording of the events that may happen to the 'population at risk'. The General Register Office was set up in 1837 and from then on was supposed to keep accurate returns of all births, deaths and marriages occurring in the country. Dr William Farr was the first holder of the post which is nowadays called 'Chief Medical Statistician' at the General Register Office.

Farr's work cannot be evaluated here, his 'Letters' published in the successive Annual Reports of the Registar General for the next forty years must be allowed to speak for themselves. What does concern us was his long continued insistence on a rational and agreed system of classification of diseases (Registrar General, 1839). As a result of the work of William Farr and his counterparts in other countries, we now have such an agreed international system kept under constant review and revision.

In Farr's day it was usual to regard cancer as one disease, although it might manifest itself in different parts of the body. Many experimentalists still seem to hold this view, but from the standpoint of the clinician and epidemiologist it is much more rewarding to think in terms of 'the cancers', a group of diseases with some common features but whose aetiology and treatment may differ widely.

EARLY RESULTS FROM STATIC DEMOGRAPHY

In the early 1900's Dr Bashford (1907), then the Director of the Imperial Cancer Research Fund, devised a classification of cancer by site that has served as a model ever since, for although many changes have been made in accordance with advancing knowledge, it has proved possible to regroup the modern subdivisions to conform to the earlier large groups, and so to keep serial records of the mortality at the major sites. At this time much concern was felt about the rising death rate from cancer, both in this country and in others where serial records had been kept. The experience as shown by the first 55 years of our General Register Office was summarized (Registrar General, 1907) as an age-standardized death rate based on the Census population of 1901, and I have extended this chart to cover the period 1911–60. Fig. 1 shows that for females the rate continued to rise until about 1920, but has since declined, whereas for males the rate has continued to rise steadily. Were it not for the continuing rise of the mortality rate from lung cancer, both

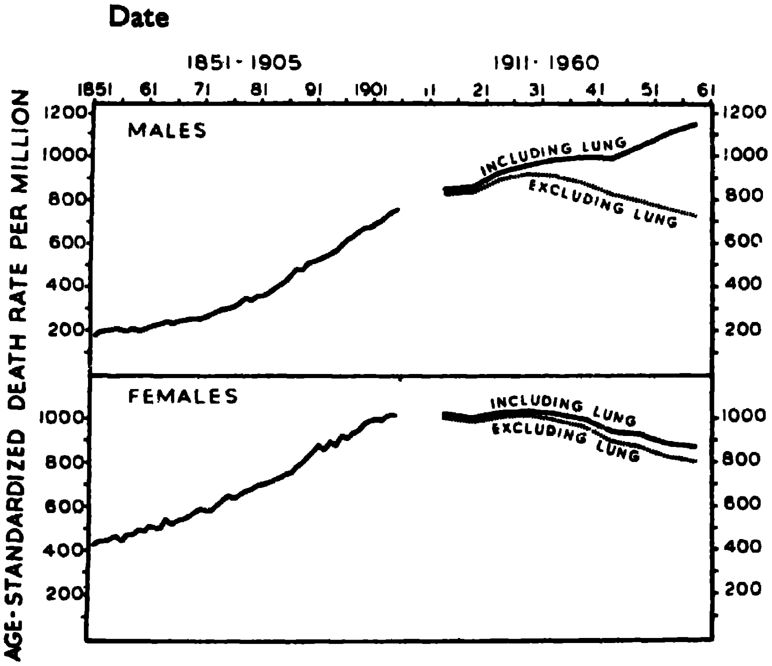


FIG. 1. Age-standardized death rates from cancer of all forms. England and Wales, 1851 to 1960. From 1911 the contribution made by cancer of the lungs is shown. (1851 to 1905 from Registrar General, 1907; 1911 to 1960 from data of Case and Pearson, 1957, and Case, unpublished.)

sexes would have benefited from a considerable decline in the total cancer death rate.

In 1915 Frederick Hoffman published his classic work *The Mortality from Cancer throughout the World.* He established beyond reasonable doubt that cancer mortality varied from country to country, from city to city both in the same country and from country to country; and that the variation in the mortality from cancer at various sites in different countries did not always follow the same pattern.

Now followed a period of consolidation and the systematic use of the accumulated records. The Registrar General in his successive supplements on Occupational Mortality (Registrar General, 1927, 1938, 1958) brought together the 'populations at risk' in different trades and professions as revealed by the census returns, and the deaths from various diseases in men (and women) who were recorded on their death certificates as being recently employed in, or retired from, these categories.

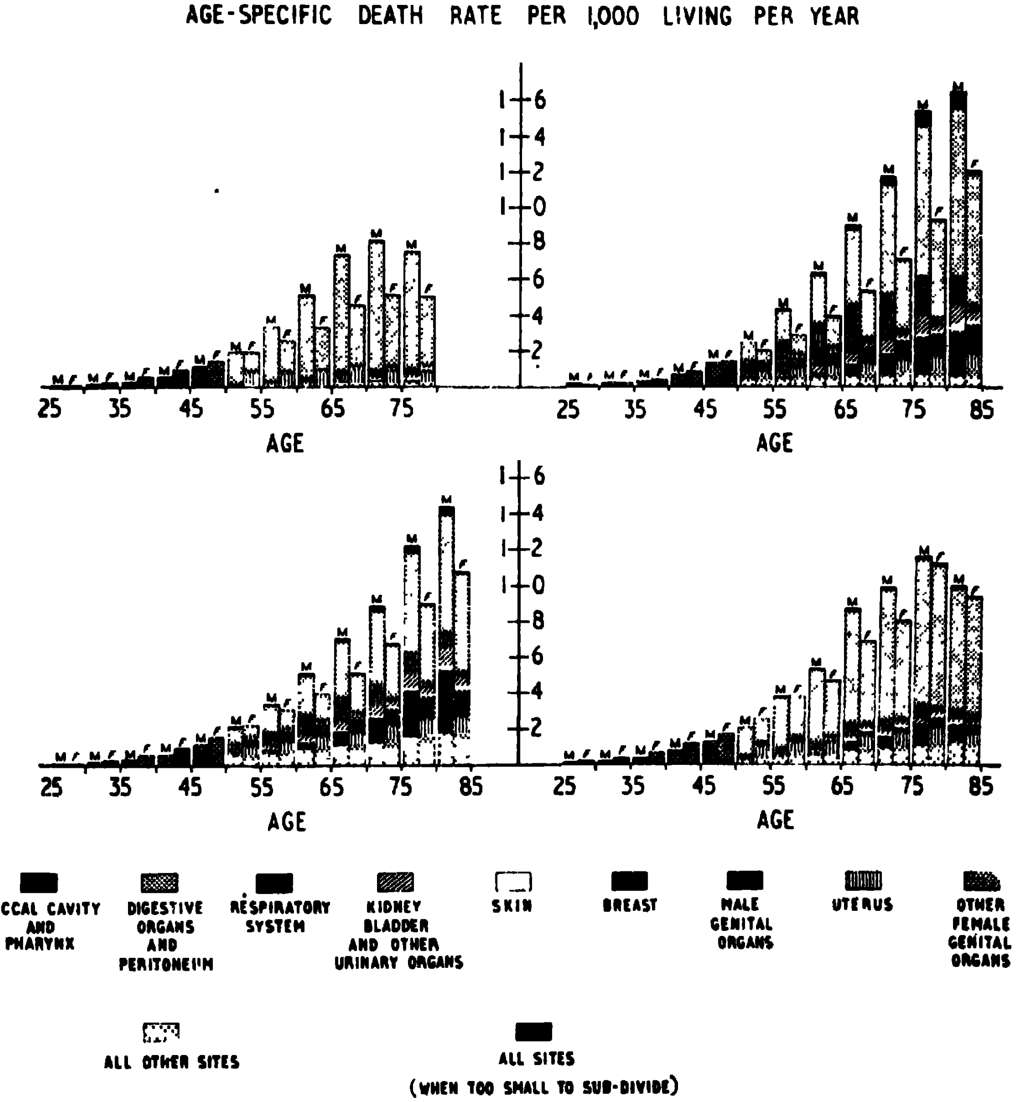
Young and Russell (1926), the Kennaways (Henry, Kennaway and Kennaway, 1931), and Henry (1956) pursued researches into occupational carcinogenesis by similar methods, and a little later Stocks (1936, 1937, 1939, 1947) began a systematic exploration of social class mortality differences and geographical variations within the country which continues to this day.

COHORT STUDIES: MORTALITY PATTERNS

In the years after the Second World War the concept of what McNalty (1961) called dynamic demography permeated into the field of cancer research. This was at least partly due to a more biological approach to vital statistics than had been shown by the actuaries, whose methods had tended to dominate medical thought. To some extent this rethinking was due, in this country at least, to the influence of those two remarkable men, J. B. S. Haldane and Lancelot Hogben.

This new approach through human biology meant that demographic studies branched into several channels, so from here on we shall have to abandon a purely chronological sequence in order to follow the developments which took place.

The dual role of nature and nurture in the determination of mortality patterns was becoming more and more appreciated, and a consideration of the principles of cohort analysis, which made allowance for this interplay, gave rise to some dissatisfaction with age-standardized rates as the sole means of expressing either time trends or international comparisons, for although they could reveal that differences of mortality did exist, and were not artifacts arising from different population structures,



Top left: Japan *Top right:* England and Wales
Bottom left: U.S.A. (Whites) *Bottom right:* Chile

FIG. 2. Age-specific death rate patterns for cancer of all forms, and for cancer at different sites, in four countries, 1951 to 1953. (Adapted from Case and Harley, 1958.)

they could not give any clear idea how these differences might have arisen (Case, 1958).

Such considerations led to the use of age-specific mortality patterns (e.g. Case and Harley, 1958) as an auxiliary to the age-standardization method (e.g. Segi, 1960) for mapping international variations in cancer mortality. Fig. 2 shows how the pattern can vary from country to country. In three of the four countries considered, the overall cancer mortality is similar, but the mortality from the different cancers shows considerable diversity. In the fourth country, Japan, the total cancer mortality is much lower than in the other three, but the mortality from the major form of cancer in Japan, cancer of the digestive organs and peritoneum, does not differ greatly from the mortality at this site in the other countries.

Another, and early, use of cohort analysis in cancer research was in the examination of the shape of the age-specific death rate pattern for cancer of the lung. Theories about a 'cancer-age' had become current, partly because of a misconception arising from some work of Dawson (1932) on breast cancer. It was thought by many that cancer of the lung was an example of a cancer which exhibited a 'cancer-age' phenomenon. However, Korteweg (1953), Levin (1953) and Kennaway and Waller (1953) showed that the peculiar pattern was merely the expression of a rapidly rising mortality rate. Formal mathematical definition of the conditions necessary for such a pattern were later given in Bignall's monograph on cancer of the lung (Case, 1958).

Fig. 3 shows the mortality rates for lung cancer amongst males in England and Wales from 1911 to 1954 (from Case, 1958). The rates are drawn in the conventional manner and also by cohorts. The illusion of a 'cancer-age' is well shown, but the cohort element of the diagram dispels this impression. Fig. 3 also shows that the rate of increase from cohort to cohort has now almost ceased at ages under 50–55 years of age, and it may well be that the rates will now stabilize at about those of the last cohort shown. Springett (1965) has recently commented on this phenomenon in England and Wales, and Alwyn Smith (1965) has pointed out that the position is similar in Scotland. These findings agree well with what would be expected if the first cohort of males to smoke 'to saturation' was that born about 1900, whose members learnt the habit in the trenches and the factories in the first World War.

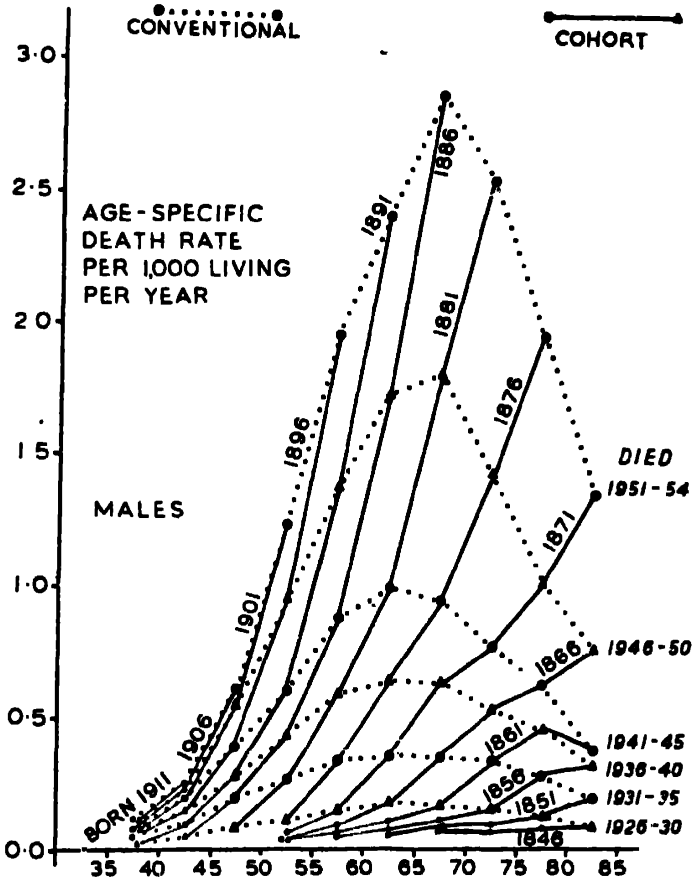


FIG. 3. Age-specific death rates from lung cancer amongst males in England and Wales, 1911 to 1954. The rates are arranged in the conventional manner and also for cohorts. (Adapted from Case, 1958.)

One result of the interest in cohort analysis was the systematic compilation of cancer death rates by site (Case and Pearson, 1957; Stocks, 1958). Tables of this sort have been used to study cohort mortality patterns and to provide a baseline for the analysis of occupational or other environmental risks in 'captive populations'.

To take the analysis of mortality patterns first, a survey (Case, 1956) of the mortality in England and Wales of the cancers by sex and some thirty sites (comprising between them over 95 per cent of all forms of cancer) showed that only cancer of the lung, tumours of the brain and other parts of the central nervous system, leukaemia and Hodgkin's disease; all for both sexes; cancer of the urinary tract in males, cancer of the pancreas in males and cancer of the ovaries and Fallopian tubes in females showed a rising mortality rate. Cancer at most other sites showed a falling mortality rate. In most sites analysed in this way, the mortality rate increases throughout life. At a few sites, such as the brain and uterus, the death rate either declines or the curve flattens with advancing years. These observations should pose interesting biological questions which have not yet been answered.

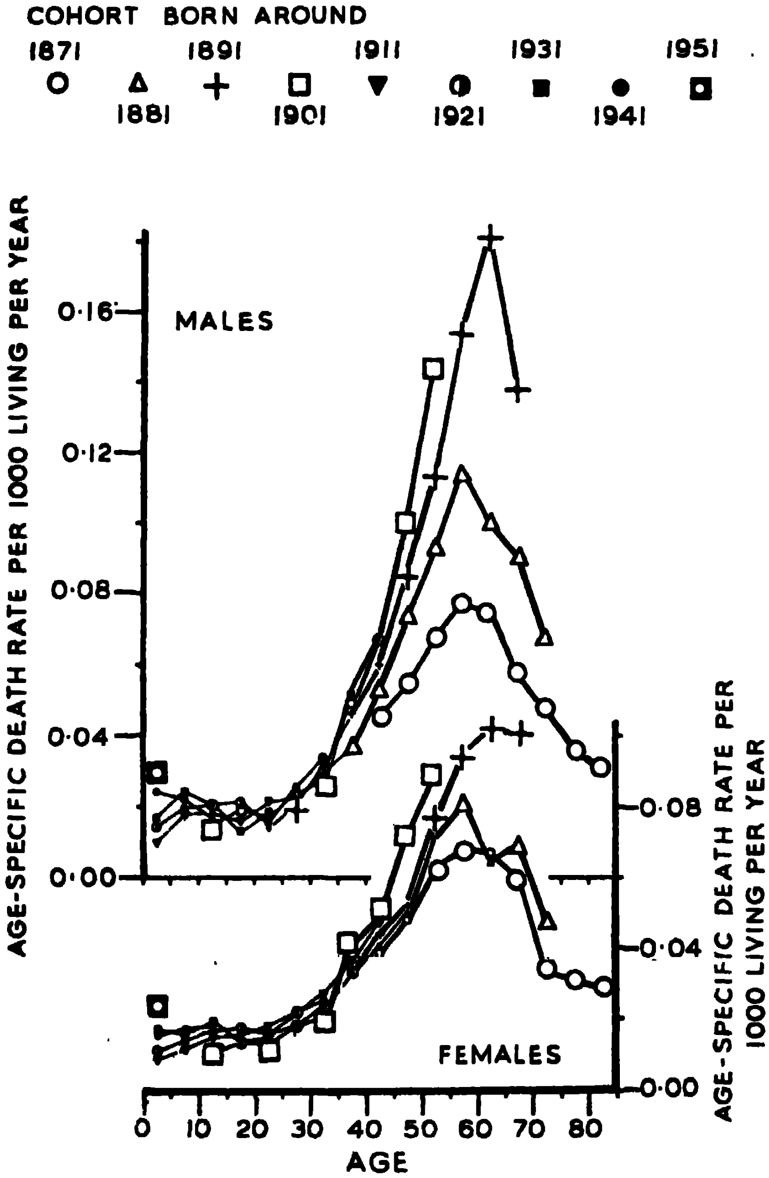


FIG. 4. Age-specific death rates from tumours of the brain and other parts of the central nervous system, England and Wales, 1911 to 1960, arranged for cohorts. (Case, unpublished.)

Fig. 4 shows this true 'cancer-age' pattern in the mortality from tumours of the brain and other parts of the central nervous system for both sexes in England and Wales during the period 1911 to 1960.

COHORT STUDIES: ENVIRONMENTAL HAZARDS

The study of environmental hazards by cohort techniques led to the development of the method, often called the prospective method, of defining a population or sub-population and then following the life experience of the survivors and recording how death overtook them. One extremely well-known example of this was Doll and Bradford Hill's (1954, 1956, 1964) continuing study on the mortality amongst doctors, subdivided according to their smoking habits. Doctors were chosen for this purpose because (*a*) they are listed during life on a special statutory register and (*b*) are likely to be distinguished after death by having their status recorded on their death certificate. This is an elementary form of record linkage, a topic to which I shall revert later.

This study has provided evidence, acceptable to all but the most bigoted, that cancer of the lung is related to the habit of cigarette smoking, and that if this habit is abandoned, the death rate of the survivors of the smoking group slowly reverts to that of their non-smoking colleagues, both in respect of lung cancer and of other causes.

Another study (Case, Hosker, McDonald and Pearson, 1954), started at much the same time, not only showed that certain dyestuff intermediates were a potent cause of cancer of the bladder, but enabled some points of human biology to be elucidated. A somewhat similar study (Case and Lea, 1955) had shown that bronchitis, either induced by mustard gas poisoning in the first World War or occurring in serving soldiers of that war for quite other reasons, was itself closely related to the subsequent development of lung cancer.

One of the points of human biology referred to was this: How serious really is the industrial risk attached to the dyestuff industry? Because in a working population there is usually a flow of new recruits and a gradual withdrawal of old workers, it is not always easy to visualize the risk to each individual worker when the disease under consideration takes many years to develop. In the dyestuff workers sufficient information was obtained to allow an assessment to be made.

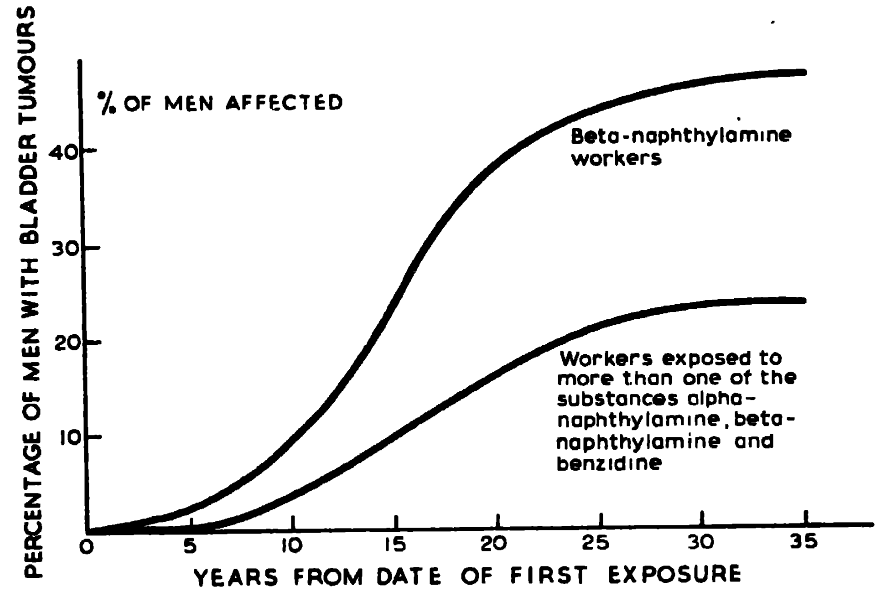


FIG. 5. To show how the incidence of occupational tumours of the urinary bladder builds up in workers exposed to certain carcinogenic aromatic amines. (Adapted from Case *et al.*, 1954.)

Fig. 5 shows how the incidence of bladder tumours in two groups of dyestuff workers gradually built up as the years passed from the time when they were first exposed to risk. In one group, who worked in a variety of processes where they would be exposed to at least two of the substances alpha-naphthylamine, beta-naphthylamine and benzidine, about 25 per cent of the men were eventually affected. In another group who were exposed only to beta-naphthylamine (of the three substances discussed) just over 50 per cent of the men contracted tumours.

COHORT STUDIES: CONCEPT OF 'IMMUNE CLASS'

In a small group of 15 men who actually distilled beta-naphthylamine, every man developed a tumour. This point is of importance because, although it is quite obvious that there is a considerable variation in individual susceptibility to a carcinogenic

TABLE 1. The 'amount of cancer' in two closed populations

Number of deaths	Cancer in dyestuff workers (Data of Case *et al.* 1954)			Cancer in chronic bronchitics (Data of Case and Lea 1955 and Case and Lea unpublished)					
				1930 to 1952			1930 to 1962		
	Cancer at all sites	Cancer of the urinary bladder	Cancer at other sites	Cancer at all sites	Cancer of the lungs	Cancer at other sites	Cancer at all sites	Cancer of the lungs	Cancer at other sites
Expected at national rates	67	5	62	156	29	127	259	62	197
Found	151+*	130	21+*	187	58	125	312	116	196

* The deaths from cancer at other sites than the bladder were incidental findings. A full investigation of all deaths was not possible.

stimulus, it had not been previously demonstrated that, in humans, there was no 'immune class'.

Certain ideas had become common in some fields of cancer research, based on poor evidence, that in any human community there is a certain 'amount of cancer', and that if circumstances increase, or decrease, the amount of cancer at one site, the amounts at other sites will decrease or increase to compensate for this (Cramer, 1936; Blond, 1954, 1955). The only rational explanation for such a phenomenon would be an 'immune class', all 'cancer susceptibles' being selected from the population by the sum total of carcinogenic stimuli, and the totals of these two classes remaining more or less constant. However, I know of no case where a captive population has been studied where such a phenomenon has been demonstrated, but of several where it has not. The two studies discussed previously, that of the dyestuff workers (Case *et al.*, 1954) and that of the war-pensioned bronchitics (Case and Lea, 1955), are two cases in point. In the first study the induced bladder cancer cases far exceed the total number of all cancer cases which would be expected at national rates in the sub-population concerned, but it was not possible to determine the full extent of the mortality from cancer other than of the bladder. In the second study, the numbers of cases of cancer of kinds other than the induced form conformed exactly to the expected national pattern at each of two analyses conducted ten years apart (Case and Lea, unpublished). These results are summarized in Table 1.

COHORT STUDIES: INTERNATIONAL COMPARISONS

A further development of the interest in cohort mortality patterns is for the purpose of international comparisons. Haenszel (1958) has shown that the mortality from gastric cancer in U.S.A. white residents has decreased dramatically over the last decade or so. Harley (unpublished) has explored the differences in the rate of this decrease in the U.S.A., Scotland, and England and Wales. Some of her results are shown in Fig. 6. At present we do not know the reasons for this great improvement, but I think that most clinicians will agree that it cannot be in any large measure due to improvements in treatment. Currently some people write, rather vaguely, about cancer mortality coming and going in waves (e.g. Lees, 1964), and they adduce this as a form of disproof of the hypothesis that cigarette smoking is in some way responsible for the present epidemic of lung cancer. It would seem to me to be more logical to say that such waves as do exist are clear evidence of an environmental origin, albeit one which we cannot always define. It is of interest to speculate whether the decline in stomach cancer in western civilization might be due to better methods of food preservation, instead of the usual speculations as to whether all

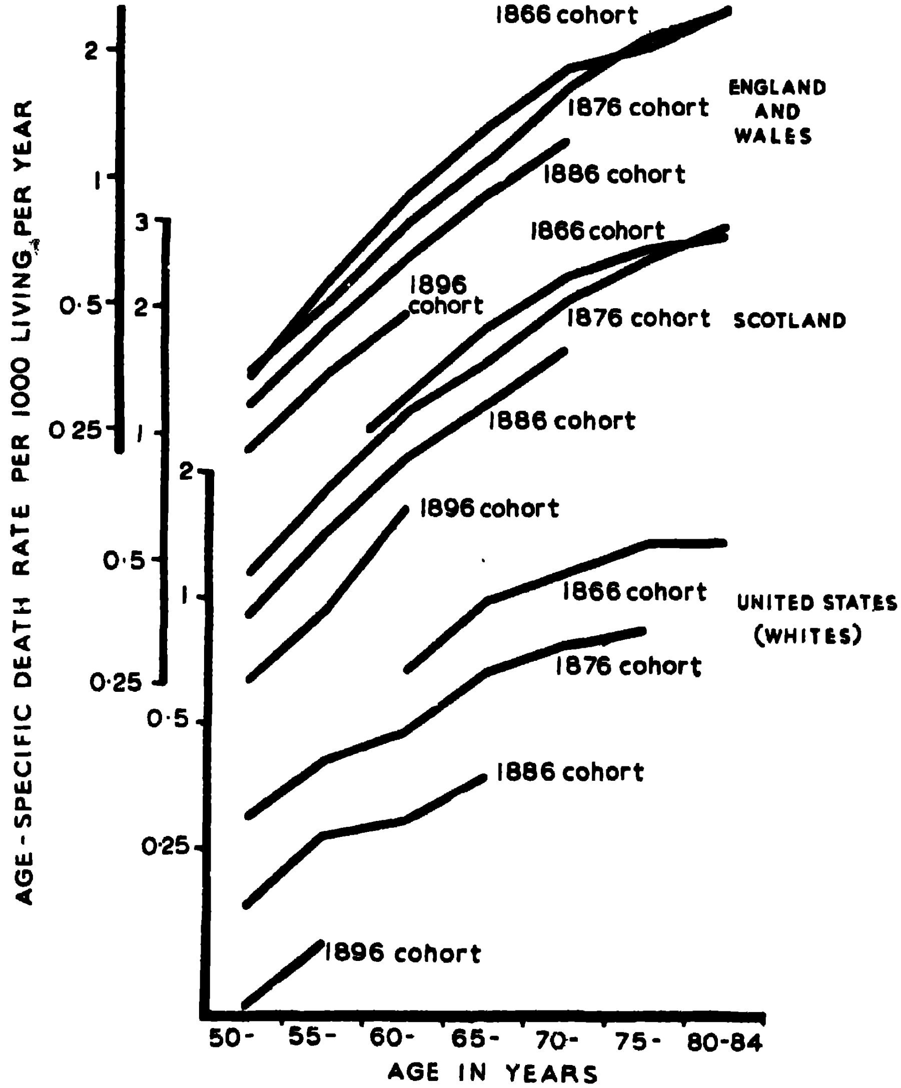


FIG. 6. Age-specific death rates from cancer of the stomach amongst females in three countries, arranged for cohorts. (Data of Haenszel, 1958; Case and Pearson, 1957, and Harley, unpublished.)

methods of preservation in themselves tend to be carcinogenic. After all, there is more reason to think that smoked food (Dungal, 1961) might be more deleterious than frozen food. The recent discovery that mould which attacks groundnuts can produce carcinogens (Leading articles, 1962, 1964) might provide a basis for suspecting that imperfectly preserved foods could contain hitherto unsuspected carcinogens of mycotic origin.

CANCER AND RACE

Other very important demographical investigations were proceeding at the same time as the investigations just discussed. In America Steiner (1954) was investigating the types of cancer occurring in first and second generation immigrants of various races, and comparing the patterns found with those prevailing in the countries of origin. His general conclusion was that environmental factors were more closely related to the development of cancer than were purely genetic ones, but the importance of extending such observations by the most refined demographic methods available can hardly be overestimated.

RECENT DEVELOPMENTS

Recently Burkitt (1963) described a peculiar form of lymphosarcoma in children in Africa. The distribution of the cases suggested that they were almost entirely confined to the climatic conditions where insect vectors could thrive. This suggested the possibility of an insect borne virus playing some part in their aetiology. Work is now being carried out to test this hypothesis.

Lea (1965) has recently noted that the mortality from breast cancer in some countries that have adequate area mortality statistics appears to be in some way latitude or temperature dependent. Although the nature of the dependence is not at present clear, studies of this sort must give rise to many fruitful lines of investigation.

All the indications are, then, that the cancers have a very considerable environmental factor in their aetiology, and that therefore preventive measures can, and must, play an ever-increasing role in cancer control as we identify more and more of these factors.

THE NEED FOR BETTER RECORD KEEPING

What of the future? As we have seen, some of the most important advances in cancer research have come from the study of 'captive populations'. This really means populations where we can follow the events that overtake the individual members as they pass through life.

To my mind, the most important steps that can be taken to ensure better understanding and prevention of cancer, a disease which at present accounts for one-fifth of our national bills of mortality, are:

(1) The immediate establishment of a national archive of all death certificates where cancer appears as contributing to death. This would enable many existing problems to be solved, and is easily accomplished. (Had such a register existed when the dyestuff workers survey was undertaken, it would have been possible to have completed the balance sheet and to have decided whether the effect of the known bladder carcinogens is in fact confined to the urinary tract.)

(2) The complete registration of all cases of cancer when they are first recognized. As the regional cancer registries now cover by far the greater part of the population, it is a pity that such registration is not made mandatory. A study of the prevalence, as opposed to the mortality, of cancer could not fail to be of value.

(3) The integration of our national records so that it is possible to proceed from a birth certificate to records of the National Health Service, those of the Ministry of Pensions, marriage certificates and the birth-certificates of offspring, until finally arriving at the death certificate.

Acheson (1964) describes a pilot scheme to study such 'record linkage'. Any suggestion that it should be applied on a national scale will certainly meet with howls of protest, for it will be claimed that any such rational system of tracing a person through life will infringe the personal liberty to indulge in such acts as evading income tax. It is worthwhile to recall that when the first proposal to take a national census was put to Parliament in 1753, the member for the city of York did not believe (Interdepartmental Committee on Social and Economic Research, 1951) 'that there was any set of men, or, indeed, any individual of the human species so abandoned as to make the proposal that we have just heard. . . . I hold this project to be totally subversive to the last remains of English liberty . . . the new bill will direct the imposition of new taxes and, indeed, the addition of a very few words will make it the most effective engine of rapacity and oppression that was ever used against an injured people . . .'

It should not, in my opinion, be impossible to devise adequate safeguards to legitimate personal liberties, whilst also providing the personal liberty to live free from the threat of avoidable disease.

At a time when most of the Ministries concerned are considering computer systems, it might be profitable for the medical profession to give its full support to such schemes as those outlined by Acheson (1964).

In the meanwhile, much can be done with piecemeal records kept for other purposes. For instance, as is now coming to light as a result of the recent publicity (Annotation, 1965) about industrial bladder cancer in the rubber and cable industries, some Trade Unions possess records that help in tracing men, or even in finding the cause of their death (Davies, 1965). The Trade Unions might seriously consider how far, both at present and in the future, their records can be used or adapted to serve as a basis for a 'built-in' prospective survey of the health of their members in different trades.

The contribution which demographic studies may be able to make to the control of cancer could well be large. What is needed is imagination and determination. The knowledge to be gained in this way can scarcely fail to become part of The Scientific Basis of Medicine of the future.

REFERENCES

Acheson, E. D. (1964) *Brit. J. prev. soc. Med.*, **18**, 8.

Annotation (1965) *Lancet*, **i**, 306.

Bashford, E. F. (1907) Table in 'Sixty Eighth Annual Report of the Registrar General of Births, Deaths and Marriages in England and Wales. 1905', pp. cii–cv. London: H.M.S.O.

Blond, K. (1954) *Brit. med. J.*, **ii**, 304 and 752.

Blond, K. (1955) *The Liver and Cancer: A new Cancer Theory*. Bristol: Wright.

Burkitt, D. (1963) *Int. Rev. exp. Path.*, **2**, 103.

Case, R. A. M. (1956) *Brit. J. prev. soc. Med.*, **10**, 172.

Case, R. A. M. (1958) In *Carcinoma of the Lung*, ed. J. R. Bignall, Neoplastic Disease at Various Sites Monograph Series. No. 1. London and Edinburgh: E. and S. Livingstone.

Case, R. A. M. and Harley, Joyce L. (1958) *Cancer Death Rates by Site, Age and Sex: 21 Selected Countries, 1951–1953*. London: Institute of Cancer Research; Royal Cancer Hospital.

Case, R. A. M., Hosker, Marjorie E., McDonald, Drever B. and Pearson, Joan T. (1954) *Brit. J. industr. Med.*, **11**, 75.

Case, R. A. M. and Lea, A. J. (1955) *Brit. J. prev. soc. Med.*, **9**, 62.

Case, R. A. M. and Pearson, Joan T. (1957) In *Cancer Statistics for England and Wales, 1901–1955*. Studies on Medical and Population Subjects No. 13. London: H.M.S.O.

Cramer, W. (1936) 'Reports of the Second International Congress of Scientific and Social Campaign against Cancer'. Brussells: Ligue National Belge Contre le Cancer.

Davies, Joan M. (1965) *AScW Journal*, **11**, 15.

Dawson, Edith K. (1932) *Lancet*, **i**, 1251.

Doll, R. (1965) *Brit. med. J.*, **i**, 471.

Doll, R. and Hill, A. B. (1954) *Brit. med. J.*, **i**, 1451.

Doll, R. and Hill, A. B. (1956) *Brit. med. J.*, **ii**, 1071.

Doll, R. and Hill, A. B. (1964) *Brit. med. J.*, **i**, 1399, 1460.

Dungal, N. (1961) *Acta Un. int. Cancr.*, **17**, 365.

Graunt, J. (1662) *Natural and Political Observations . . . upon the Bills of Mortality*. London: Martin, Allestry and Dicas.

Haenszel, W. (1958) *J. nat. Cancer Inst.*, **21**, 213.

Henry, S. A. (1956) *Cancer of the Scrotum in Relation to Occupation*. London: Oxford University Press.

Henry, S. A., Kennaway, Nina M. and Kennaway, E. L. (1931) *J. Hyg., Camb.*, **31**, 125.

Hoffman, F. L. (1915) *The Mortality from Cancer throughout the World*. Neward: N.J., The Prudential Press.

Interdepartmental Committee on Social and Economic Research (1951) 'Census Reports of Great Britain, 1801–1931'. Guides to Official Sources No. 2. London: H.M.S.O.

Kennaway, E. L. and Waller, R. E. (1953) *Acta Un. int. Cancr.*, **9**, 486.

KORTEWEG, R. (1953) *Acta Un. int. Cancr.*, **9**, 529.

LEA, A. J. (1965) *Brit. med. J.*, **i**, 488.

LEADING ARTICLE (1962). (1964). *Brit. med. J.*, **i**, 309, and **ii**, 204, respectively.

LEES, T. W. (1964) *Lancet*, **i**, 1393.

LEVIN, M. L. (1953) *Acta Un. int. Cancr.*, **9**, 531.

McNALTY, A. S. (Ed.) (1961) *The British Medical Dictionary*. London: The Caxton Publishing Co.

REGISTRAR GENERAL (1839) 'First Annual Report of the Registrar General of Births, Deaths and Marriages in England', pp. 86–166. London: H.M.S.O.

REGISTRAR GENERAL (1907) 'Supplement to the Sixty Fifth Annual Report of the Registrar General of Births, Deaths and Marriages in England and Wales. 1891–1900'. London: H.M.S.O.

REGISTRAR GENERAL (1927, 1938, 1958) 'The Registrar General's Decennial Supplement, England and Wales, 1921, 1931 and 1951'. Occupational Mortality volume. London: H.M.S.O.

SEGI, M. (1960) *Cancer Mortality for Selected Sites in 24 Countries (1950–1957)*. Sendai: Tohoku University School of Medicine; Department of Public Health.

SMITH, A. (1965) In 'Supplement of the Annual Report of the Registrar General for Scotland 1963'.

SPRINGETT, V. H. (1965) Meeting of the Thoracic Society. London. February.

STEINER, P. E. (1954) *Cancer: Race and Geography*. Baltimore: The Williams and Wilkins Company.

STOCKS, P. (1936, 1937 and 1939) 'British Empire Cancer Campaign Annual Reports Nos. 13, 14 and 16', pp. 239, 198 and 308, respectively.

STOCKS, P. (1947) *Regional and Local Differences in Cancer Death Rates*, Studies on Medical and Population Subjects No. 1. London: H.M.S.O.

STOCKS, P. (1958) In *Cancer*, ed. R. W. Raven, Vol. 3, pp. 184 *et seq.* London: Butterworth.

YOUNG, M. and RUSSELL, W. T. (1926) *An investigation into the Statistics of Cancer in Different Trades and Professions*. M.R.C. Special Report Series No. 99. London: H.M.S.O.

VI

The Age of Reaching Puberty: Some Associated Factors and Some Educational Implications

J. W. B. DOUGLAS

INTRODUCTION

IMPORTANT educational problems are raised by the fact that there is a wide range of ages at which children become sexually mature, for boys the majority reach puberty between 10 and 18 years, while for girls the parallel ages are 8 and 17 years. In the National Survey which I will shortly describe, for example, the earliest reported age at menarche was 8 years.

This variation in the age of reaching puberty means that for a large part of the school years each age group of children will have members at very different stages of sexual maturity. In the primary schools some girls and a very few boys will show signs of puberty, while at the other extreme, in the sixth forms a few girls and more boys will still be at an infantile or pre-adolescent stage. If the intelligence of children of the same chronological age is related to their stage of sexual maturity, then problems of secondary selection are raised. I discuss in this paper, first certain biological and social factors associated with the age of onset of puberty, and second the educational progress of early and late maturing children.

AVAILABLE INFORMATION

The data described in this paper are derived from a study of 5,362 children, all born in one week of 1946, who have been under observation since birth. In 1961 4,592 of these children were still living in Great Britain. They come from all parts of the country and from every type of home, and during the nineteen years we have kept in touch with them our losses have been light. We know the type of school attended by more than 98 per cent of those who are still living here and we lose no more than 10–11 per cent for any given examination or interview. I do not propose to go into details of how this study evolved as this information is available elsewhere (Douglas, 1964). It is called the National Survey of Health and Development and was originally sponsored by the Institute of Child Health, the Society of Medical Officers of Health and the Population Investigation Committee under the Chairmanship first of the late Professor James Young and then of Sir Alan Moncrieff.

The relevant aspects for this paper are the school doctors' special examinations of the children at 11 and 15 years and the intelligence and school achievement tests given by the teachers when the children were 8, 11 and 15 years. As the children are spread all over the country, from the Shetlands to the Scillies, it was impossible for us to examine and test them ourselves, so we asked for help from the school doctors, school nurses and teachers, help which was generously given. The school doctors gave these children special medical examinations, lasting an hour or more, on four occasions, of which two only are relevant here. At 11 they asked the mothers, who attended with their children, whether the girls had started their periods and if so the date; they described the state of development of the breasts and noted the presence or absence of pigmented pubic hair. At this age 1·4 per cent of the girls had started their periods, 15 per cent were pubescent, having enlarged breasts and pigmented pubic hair though no periods, and a further 30 per cent had some breast enlargement only; so that nearly half the girls showed some sign of advancing puberty. For the boys at this age pigmented pubic hair was reported for 3·5 per cent.

At 15 the mothers were again asked about their daughters' periods and in addition the school doctors reported on the stage of development of the breasts and the presence of pubic hair and axillary hair. For the boys the doctors also assessed the stage of development of the genitals and noted whether there were voice changes. These records by the school doctors have been used to group both the boys and the girls into four stages of puberty.

The 1,817 girls for whom date of menarche was recorded were divided into the following four groups according to the reported dates of onset of their first period:

(i) Menarche before 12 (238 girls)
(ii) Menarche in 13th year (450 girls)
(iii) Menarche in 14th year (609 girls)
(iv) Menarche at 14 or later (520 girls)

This grouping is used when looking at the association between puberty and intelligence or school performance.

The classification of the boys by the stage of puberty they had reached was based on the school doctors' description of their secondary sexual characteristics at the 15-year-old medical examination. There are four groups:

(i) Mature:	fully broken voice, axillary hair, profuse pubic hair, mature genitals	(489 boys)
(ii) Advanced:	genitals assessed as mature, but one or more of the other maturity signs absent	(599 boys)
(iii) Early signs:	intermediate between advanced and infantile	(700 boys)
(iv) Infantile:	infantile genitals, no pubic or axillary hair, voice unbroken	(207 boys)

A total of 1,995 boys were examined.

Some Biological and Social Factors Associated with the Rate of Sexual Development

Before giving details of the intelligence and attainment tests used in this study I will describe some associations and lack of associations between these maturity groups and certain social and biological characteristics of the children and their families.

There is no doubt that in the last century menarche was earlier among the daughters of the well-to-do than of the poor. Tanner (1955), for example, quotes a study of Manchester women in 1820 which showed the educated ladies as reaching menarche just over a year in advance of the working women. And Marro (1900) at the end of the nineteenth century gave distributions of menarche for Piedmontese girls which yield exactly the same difference between rich and poor. In the underdeveloped countries today there is also evidence of a similar difference in the age of maturation. It has been pointed out that in Bristol, grammar school girls reach puberty at an earlier age than secondary modern school girls, but for reasons that will be evident later this is adequately explained by what happens at secondary selection and cannot be taken as evidence of an association between poverty and delayed maturation. The National Survey, as Table 1 shows, provides no evidence of any association between the age at which children reach puberty and the social class of their parents (Douglas, 1964). A similar lack of association with age of puberty is found when the children are grouped by their father's occupation rather than by the rather more complex social classification shown in this table.

TABLE 1. Proportion of early maturing children by social class

		Social class				
		Upper Middle	Lower Middle	Upper Manual	Lower Manual	Significance P
Advanced or mature boys	%	56·9	54·8	55·1	53·0	0·8–0·7
Girls reaching menarche before 13th birthday	%	40·4	36·4	38·5	37·0	0·9–0·8

The unexpected finding that social class and puberty are not associated is supported by recent studies in Aberdeen (Nisbet and Illsley, 1963), and Bristol (Rogan, 1958), and it seems that at some time during this century economic factors ceased to exert a differential influence on the rate of sexual development of children in Great Britain. That it is not possible to be more precise than this is a sad reflection on the records of the school health service and on the hopes expressed when it was set up in 1907 that it would provide national statistics of health and growth.

That there is a relation between number of sibs and menarche is well known and is illustrated in Fig. 1 by the Survey data.

At the age of 13, 54 per cent of the girls with no brothers or sisters have reached menarche, as compared with 35 per cent of those with four or more. Put another way, half the girls had reached menarche at or before 12 years 9 months if they were only children and at or before 13 years 4 months if they were members of the larger families. Exactly the same sort of relationship holds for the boys, though it is less neatly shown because we have no continuous variable for assessing the age they reach puberty.

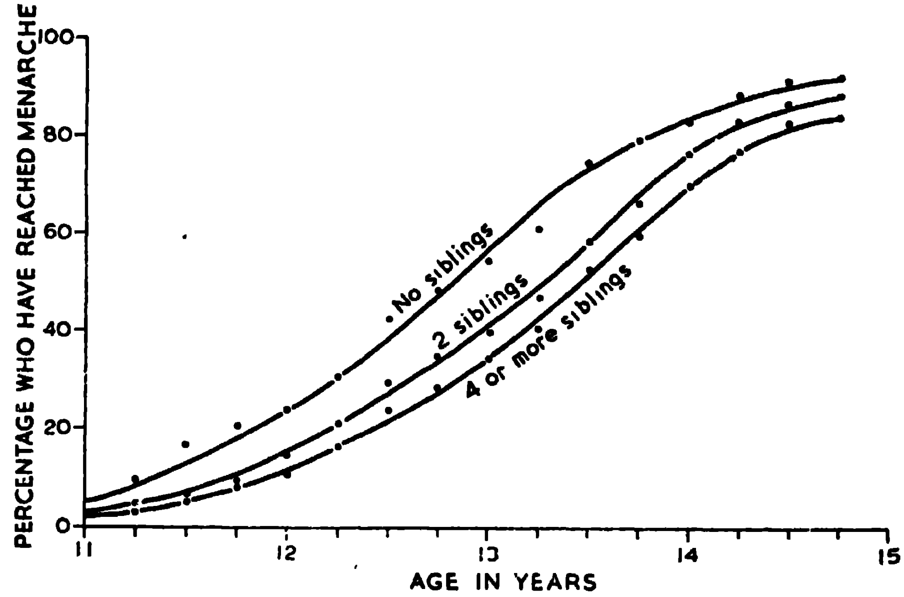


FIG. 1. Age of Menarche-children from families of different sizes.

The immediate question to ask is whether the development of first-born children is similar to that of only children or similar to that of their brothers and sisters. In other words, is it birth rank or family size that matters? The answer to this question is shown in Fig. 2 which gives the proportion of first-born boys who were assessed as mature or advanced at 15 and the proportion of first-born girls who reached menarche in their 13th year or earlier. These children are further grouped according to the number of their sibs at 15 years, by which age their families were virtually complete in size.

It is immediately clear that the sexual development of these children, both boys and girls, is related to the size of their families and not to their position in the family. There is, however, a difference between the sexes; for the boys there is a regular gradation with increasing family size, whereas for the girls the split is predominantly between the only child and the rest. As these are all first-born children the suggestion from the graph is that it is not the intra-uterine conditions that are important, but events occurring after birth, since if the former were true then all first-born should show advancement of puberty whatever their family size.

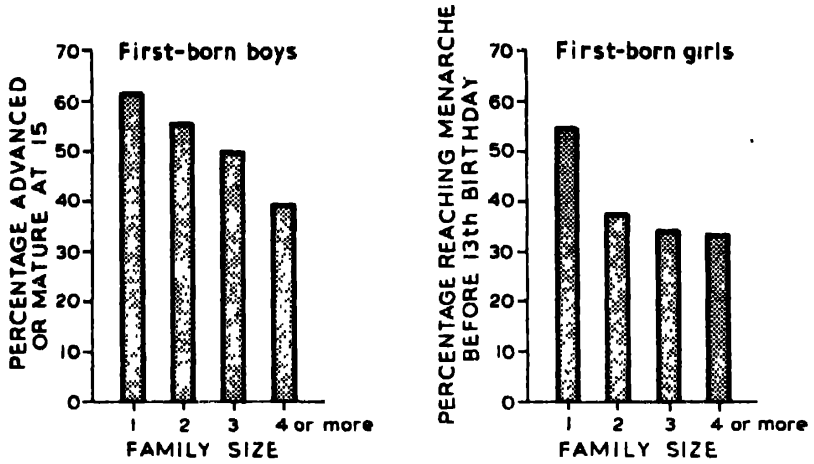


Fig. 2. Proportion of early maturing first-born children grouped by family size.

It seems that the separation of a child from his sibs by a wide birth interval is another factor connected with the age of onset of puberty, as will be seen in Fig. 3. This compares the distribution of menarche for girls in two-child families, who are separated from their elder sib by an interval of 3 years or less with those separated by more than 6 years. Those born after the longer interval reach menarche at a considerably earlier age than those born after the shorter interval; 59 per cent of the former and 42 per cent of the latter have reached it by thirteen. If the importance of birth spacing is confirmed it will suggest that the early social experiences of children are related to the advancement or postponement of puberty.

Nutritional differences are unlikely to explain the association of early puberty with small family size, as this association is equally marked in each social class. For the girls, moreover, the main difference is between only children and the rest, and there is no evidence, in the middle classes at any rate, that the younger children in families of two are today under-nourished.

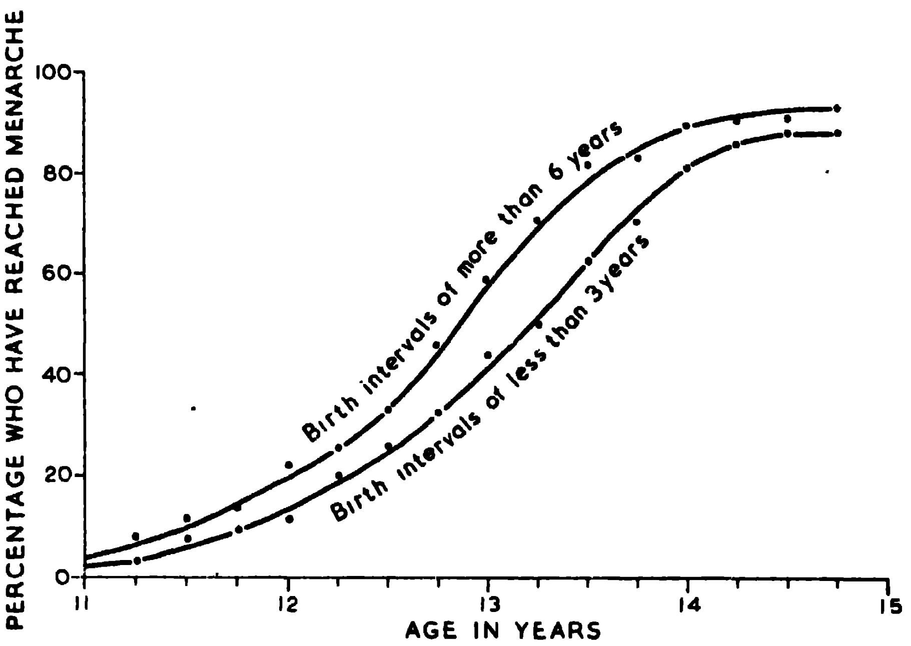


Fig. 3. Age of Menarche—younger children in two-child families grouped by birth interval.

Differences in breast feeding might, of course, be involved, but there is no evidence of this in the National Survey as will be seen in Table 2.

Table 2. Early and late maturing children compared for history of breast feeding and weaning and for parents' ages

		Boys Signs at 15 Mature or Advanced	 Infantile or Early	Girls Menarche Before 13th birthday	 13th birthday or later
Breast fed for less than one month	(%)	35·0	37·8	32·9	33·5
Weaned from breast and bottle at 10 months	(%)	62·0	60·0	65·3	61·8
Average age of mother at child's birth	(yrs)	29·0	29·0	29·0	29·2
Average age of father at child's birth	(yrs)	32·1	32·0	32·0	32·2

Another explanation of the association between puberty and family size might be that in a study such as this, which is based on a single week's births, giving as it were a sample of the reproductive performance of all fertile women, the parents of small families are likely to be younger than the parents of large. And if the trend towards early puberty is a generation effect, that is to say, if the children of each generation of parents have their own pattern of maturation, those from the older generation reaching maturity at a later age than those from the younger, this might explain some of the family size differences already noted. But in fact there is no evidence of this. The age of both the fathers and the mothers of these children does not vary between the maturity groups; for example the mothers of the girls who reached menarche in their 13th year or earlier were, on the average, 29 years old exactly at their daughter's birth, whereas the mothers of girls who matured later than this were 29 years 2 months (see Table 2).

Children from small families get less infection during their early years than children from large families, and a heavy load of illness in the early life of the latter might be expected to lead to a disturbance of growth and perhaps of sexual maturation. But there is no evidence of this in the Survey. Time spent in hospital during the pre-school years is as great among the early maturing children as among the late, and within each size of family the early and the late maturers are equally likely to get colds and lower respiratory infections.

So far we have been unsuccessful in identifying any social factor or complex of factors that are associated with early or late puberty. As Table 3 shows, there is no consistent relationship between maternal care and puberty. Among the boys favourable assessments of the mother's care and understanding of her child are rather more often reported for the early maturers; among the girls the same trend is found, though it is less marked, and for neither sex is it statistically significant when only children are considered so as to avoid the effects of family size. There is a slight suggestion in two-child families that girls who have brothers mature earlier than girls who have sisters, and that boys who have sisters mature earlier than boys who have brothers, but again these differences are small and statistically not significant. At later ages we find no evidence that children in mixed sex classes at either primary or secondary schools are advanced in puberty as compared with those in single sex classes. The slight excess of early maturing children at single sex secondary schools is explained by the higher proportion reaching grammar school.

TABLE 3. Early and late maturing children compared for maternal care, sex of sibs and sex of class

			Boys Signs at 15		Girls Menarche at	
			Mature or Advanced	Infantile or Early	13th year or earlier	14th year or later
ONLY CHILDREN						
% of mothers most favourably assessed by H.V's for	Care and understanding of child	%	46·2	37·5	46·1	44·0
	Cleanliness of child and home	%	45·5	47·6	52·0	51·0
CHILDREN IN 2-CHILD FAMILIES						
Younger sib opposite sex		%	53·8	51·8	53·8	46·7
Elder sib opposite sex		%	50·0	49·9	55·5	52·6
ALL CHILDREN						
In single sex Primary school classes		%	13·6	11·6	13·2	13·2
In single sex Secondary school classes		%	54·4	51·3	56·7	53·8

The observations so far suggest that the rate of sexual maturation is decided at a point after the birth of the child and before the age of 7. This upper limit of 7 is suggested by the fact that at 7 boys and girls who will mature early are already taller than those who will mature late, even after allowances have been made for the size of family from which they come, whereas at birth they weighed the same (Douglas and Simpson, 1964). It seems probable that the rate is partly set by the isolation or non-isolation of the child within his family, by the nature of his relations with his parents and perhaps with his sibs.

PUBERTY AND EDUCATIONAL PERFORMANCE

The boys and girls in this study completed three sets or batteries of educational tests at 8, 11 and 15 years. These tests were chosen for us by the National Foundation for Educational Research in England and Wales, who also scored them and then standardized the results. They were given to the children in the majority of instances by the teachers. Each battery contained tests of attainment as well as of intelligence. At 15, for example, the tests used were of reading, mathematics, verbal intelligence and non-verbal intelligence. These tests were each standardized on the whole group of pupils taking the tests and the resulting 'T scores' were given a mean of 50 and a standard deviation of 10. There is of course no difficulty in transforming these into 'IQ's' with a mean of 100 and a standard deviation of 15, but for a variety of reasons we prefer to use T scores. For each year there is also an average score for all tests which was obtained by adding together the individual T scores and then re-standardizing this aggregate score on the group of children who had taken all three batteries of tests, i.e. at 8, 11 and 15 years. The correlations between these aggregate scores are 0·82 at 8–11, 0·87 at 11–15 and 0·79 at 8–15.

In addition to these tests I refer to the results in 'O' Levels. A 'good' 'O' Level certificate is defined as one which has at least four passes in at least three of the following fields: English language, mathematics, a natural science and a foreign language (excluding Welsh) gained at or before $16\frac{1}{4}$ years.

The average aggregate scores at 8, 11 and 15 for the boys and girls grouped by their level of maturity at 15 years are shown in Fig. 4.

The scores at each age refer to the same children, and of course each child stays throughout in the same maturity group. There is a relatively constant difference between the extreme groups throughout the seven years. At 8 years, and well before any of the boys had shown signs of puberty, those who were later to be assessed as mature at 15 scored 4·8 points of T score more than those who were later assessed as infantile. And at 15 the gap between them was 4·1 points. Similarly the girls who reached menarche before 12 scored in the 8-year test an average of 3·2 points more than those who reached menarche after 14 years, and in the 15-year tests they scored 3·7 points more. There is no suggestion here of a spurt in intelligence or attainment to match the prepubertal spurt in physical growth. The intermediate groups follow a rather puzzling pattern. At 8 the 'advanced' boys score exactly the same as those with 'early signs'; during the next seven years the two groups diverge, so that by 15 there is a difference of 1·6 points in favour of the more mature group. Similarly, the girls reaching menarche between 12 and 13 make exactly the same scores at 8 years as the girls reaching menarche between 13 and 14, whereas at 15 they are 1·4 points ahead.

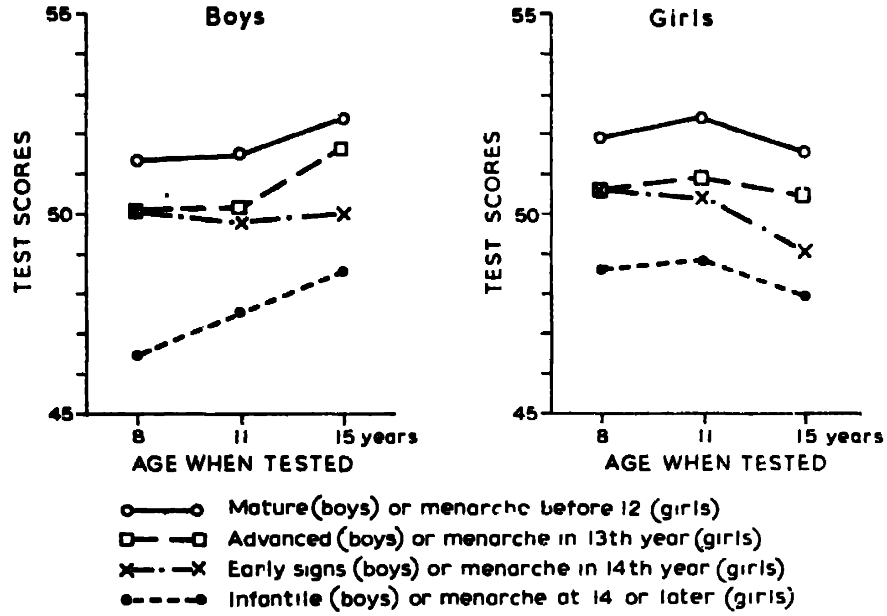


FIG. 4. Mean aggregate test scores of boys and girls grouped by stage of puberty.

Fig. 4 also shows, between 8 and 15 years, a change in the relation between the scores made by boys and girls. At 8 and 11 the girls make consistently higher scores than the boys, but at 15 they do considerably worse. This relative deterioration of the girls will be discussed in a later publication.

There appears to be no inequality in the allocation of grammar school places to the early or late maturing children after allowing for the differences in measured ability (Douglas, 1964). The fact that there is no tendency for the late maturing children to improve their test performance relative to those maturing earlier, up to the age of 15 at any rate, suggests that in the secondary selection examinations there is no need to make an allowance for maturity. With the extreme maturity groups there is no evidence that the late maturers catch up, and in the intermediate groups the earlier maturing children actually increase their advantage between 11 and 15 years.

In the later years at secondary school the early maturing children continue to have an advantage over the late maturers. They get more good 'O' Levels in the General Certificate of Education and they are more likely to stay on for a sixth year at school.

In general it seems that the early maturing children retain their initial advantage until the end of their years at secondary school. We have not yet had the opportunity of seeing whether they retain it at the university—some studies suggest that they do not, perhaps owing to selection by marriage or by a desire to achieve independence. We do find (Douglas and Ross, 1964) that early maturing manual working class girls are more likely to leave grammar schools at the end of the fifth year than late maturing manual working class girls, but this, we suspect, may be rather that they are more conscious of the conflict between the values of home and school and the need to conform to the pattern of the *home* adolescent group, which is in favour of early leaving, than that they deteriorate in ability.

The obvious question to ask is whether the differences in test performance described above are explained by the relation between puberty and family size. It is well known that as the size of families increases so the average measured ability of the children declines, and the question is whether the puberty effect is partly or wholly explained by the family size effect or whether part or the whole of the family size effect can be attributed to puberty differences. In *The Home and the School* it was shown that the test scores of 11-year-old girls with no brothers or sisters were similar whether they were early or late maturers. This can now be confirmed for boys as well as girls and for test performances at all three ages. The results are given in Fig. 5.

There is no suggestion that among only children early puberty is associated with good test performance and late puberty with poor. This is in striking contrast with the results for the whole sample (see Fig. 4) which show the early maturing children as having a consistent advantage over the late maturing ones. Among the only children the infantile boys make rather lower scores than the rest, but there are only 26 of them and with this small number their poor performance could well be explained by chance selection. The only girls reaching menarche in their 13th year make low scores which at 8 years, but not at 11 or 15, are significantly below those of the rest of the girls. No reason can be advanced for this peculiarity.

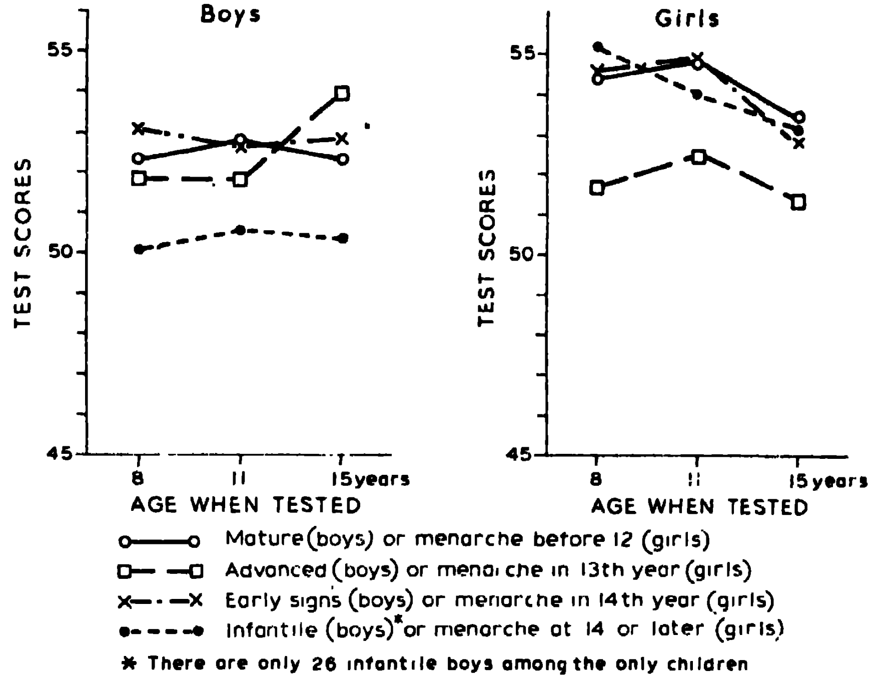


FIG. 5. Mean aggregate test scores of only children grouped by sex and stage of puberty.

The position is less clear with children from the larger families probably because there are many complicating factors of position in family and birth spacing to take into account. In each size of family taken separately the difference in scores of the early and late maturers still persists but is less than when the sample is taken as a whole. I consider that the only children provide the critical case. If there is no association between stage of puberty and measured intelligence in only children, then it is unlikely that there is a meaningful association between these two variates in the larger families.

SUMMARY

The conclusions then are very simple. Today the age of puberty of both boys and girls is similar in all social classes, but it differs considerably with family size, being advanced in the small families and retarded in the large. For the girls, the only children in particular stand out as reaching menarche early. The fact that the association between puberty and family size is as marked for the first-born children as for the later born points to events occurring after birth as helping to set the pace of sexual development, and a study of birth intervals and heights suggests that the first seven years are the crucial ones.

In the schools and in our tests at 8, 11 and 15 the early maturing children do better than the late maturing ones. There is certainly no evidence that the late maturing children catch up with the early maturers in either measured ability or school performance up to the age of 16¼; there is therefore no need to make allowances for sexual maturity in secondary selection. The advantage of the early maturing children in the tests of intelligence and attainment and at school seems to be adequately explained by the association of puberty with family size.

ACKNOWLEDGEMENTS

The National Survey of Health and Development was sponsored by a joint committee of the Institute of Child Health, the Society of Medical Officers of Health and the Population Investigation Committee. It is now being directed from the Medical Research Council Unit at the London School of Economics. In its early years the inquiry was financed by the Nuffield Foundation, and grants for continuing it through the school years have come from many sources, including the Ford Foundation and the Population Council, Inc.

We have received much assistance from the National Foundation for Educational Research in England and Wales—they supplied us with all the school tests and both checked the markings and standardized the scores. We wish to thank them for all their help, and also the Scottish Council for Research in Education for advice we have had on many occasions.

A national longitudinal study such as this is only possible because

of the generous co-operation we have had from many people and agencies: here we would especially like to thank the directors of education, teachers and health visitors in all parts of Great Britain, who have given us information about the boys and girls included in this Survey. We should also like to thank the mothers who willingly answered numerous detailed questions, and the boys and girls themselves.

This study of puberty in the National Survey sample is being made jointly with Miss J. M. Ross of the Medical Research Council Unit at the London School of Economics. I wish to take this opportunity to thank her for helping me with the preparation of this review and the presentation of the results. I should also like to thank Mrs E. Wilson of the Department of Geography, London School of Economics, for preparing the diagrams.

REFERENCES

Douglas, J. W. B. (1964) *The Home and the School.* London, MacGibbon & Kee.
Douglas, J. W. B. and Ross, J. M. (1964) *J. Child Psychol. & Psychiat.*, **5**, 185.
Douglas, J. W. B. and Simpson, H. R. (1964) *Milbank Memorial Fund Quarterly*, **42**, 20.
Marro (1900) *La Puberta*, Turin. (Quoted by Niceforo, A. (1905) *Les Classes Pauvres*, Paris, Giard & Brière.)
Nisbet, J. D. and Illsley, R. (1963) *Brit. J. ed. Psychol.*, **33**, 169.
Rogan, E. (1958) Report of the Bristol Medical Officer of Health, Appendix B.
Tanner, J. M. (1955) *Growth at Adolescence.* Oxford, Blackwell.

VII

Disturbances of the Immunoglobulins

J. R. HOBBS

FOLLOWING the introduction of electrophoresis by Tiselius it became apparent that the antibody activity of human serum was largely in the fraction with γ-mobility, thus called γ-globulin. Later Waldenstrom with the aid of the ultracentrifuge established that antibody activity could be present in both the major (7S) γ-globulin, now known as γG,[1] and in a macrogloublin fraction (19S), now known as γM.[1] Then Heremans established that yet another distinct 7S globulin, now known as γA,[1] could carry specific antibody activity, and in 1959 proposed the concept of the immunoglobulins to replace that of γ-globulin. Already a fourth, γD,[1] has been described by Rowe and Fahey (1965). An immunoglobulin is a protein in which specific antibody activity can be demonstrated, though it must be admitted that, at present, known specific antibody activity can only account for some 5 per cent of the total immunoglobulins of human serum. This is probably due to our ignorance of all the antigens to which humans are exposed, for specific antibody can only be defined in terms of its antigen.

By Burnet's theory, a solitary antigen could stimulate the proliferation of a single clone of cells to produce antibody. The protein molecules produced should all have the same structure so that on electrophoresis they should all have the same mobility and run as a solitary narrow band (Fig. 1). Such monoclonal antibody has actually been observed in a few rare instances where a single hapten has been used as antigen, enabling the isolation and purification of its antibody (Singer, 1965). More often, however, even purified artificial haptens evoke several overlapping antibodies, and with time even more appear, that is one antigen can evoke a spectrum of antibodies. Pure proteins can have several antigenic sites in the one molecule and the natural challenge of an infecting organism can involve many many proteins. Such a broad spectrum of antigens results in the production of an even broader spectrum of antibodies so that following infections all we can see on electrophoresis is a diffuse polyclonal increase of immunoglobulins with mobilities which in fact stretch from α_2 to γ_4.

---

[1] See 'Nomenclature of Human Immunoglobulins' (1964), *Bull. Wld. Hlth. Org.*, 30, 447.

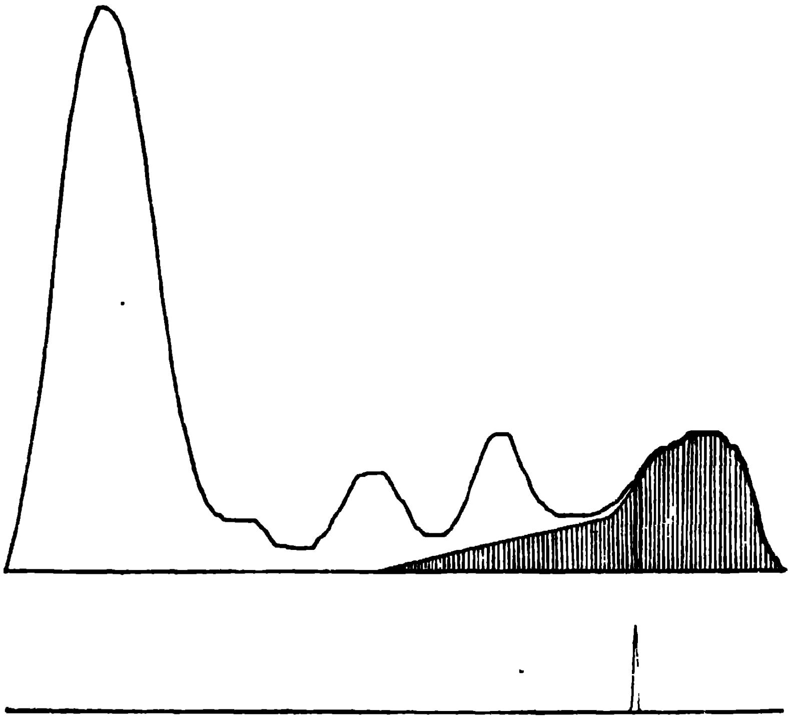


FIG. 1. The concept of monoclonal and polyclonal gammopathies. Monoclonal immunoglobulin can have a narrow homogeneous electrophoretic mobility anywhere from α_2 to γ_4. Polyclonal increases are broad and heterogeneous.

Among the disturbances of the immunoglobulins, simple electrophoresis can divide them into three main groups, the deficiencies seen as hypo-γ-globulinaemia and the excesses which may be monoclonal or polyclonal.

NORMAL IMMUNOGLOBULINS

By immunoelectrophoresis (Scheiddeger, 1955) some 20–40 proteins can be identified in human serum, and among these are our own immunoglobulins acting as antigens. These four are antigenically related to one another and the position was beautifully clarified by the fundamental work on immunoglobulin structure reviewed by Professor Porter in a lecture on the Scientific Basis of Medicine last year. Fig. 2 shows schematically that immunoglobulin structure is common throughout. Two A chains are joined to each other by three sulphydryl bonds, and then to two B chains by two more bonds. γM consists of six such units bound together by sulphydryl bonds. The B chains are common throughout, so that an antiserum made against one purified immunoglobulin usually cross-reacts with the B chains of the others. The normal B chains are physicochemically very heterogeneous but from the antigenic viewpoint are of two main

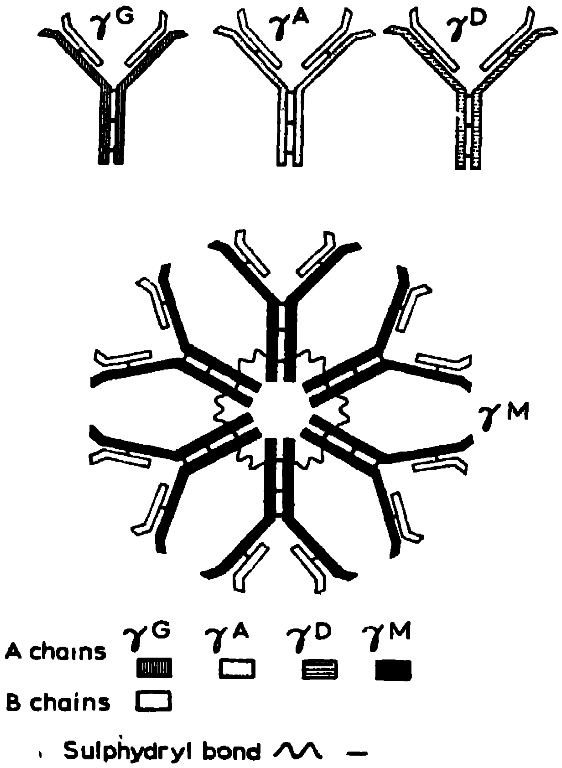


FIG. 2. A scheme of immunoglobulin structure. The notches at the extremities of the arms represent antigen-combining sites.

types, 70 per cent are type K and 30 per cent are type L. This proportion holds good throughout normal immunoglobulins, and throughout large series of myeloma globulins, however we believe individual molecules only contain one type of B chain. Thus monoclonal B chains, previously recognized as Bence-Jones proteins, are either type K or type L. With samples of each type it is possible to absorb out antisera to immunoglobulins, leaving only the activity against the A chains. The A chains carry antigenic differences which enable us to recognize the individual immunoglobulins. Such monospecific sera are used to type immunoglobulins, and can also be used to measure them using gel-diffusion methods such as that of Soothill (1962*a*), kindly taught me at Birmingham.

γG comprises 75 per cent of the total serum immunoglobulins (600–1600, mean 1100 mg/100 ml), has a mol. wt. 160,000, and a T½ of 20–30 days which is dependent on its concentration (Rogers-Brambell, 1964). Many kinds of antibody activity have been found to be γG, and as γG circulates throughout the interstitial fluid its main function seems to be to inactivate any dissemination of antigen through the body fluids.

γA, about 20 per cent of the total (mean about 300 mg/100 ml) has a similar mol. wt. but a much shorter half-life around five days. Several kinds of antibody have been identified as γA, especially reagins, so that γA seems to have a tail which will readily fix to our own tissues. It is secreted into colostrum, saliva, the bronchi and the gut. With reaginic antibody fixed to the gut wall, any antigen trying to become attached would set off a local anaphylactic reaction and this might be useful in keeping off helminths.

γM, some 5 per cent of the total (mean about 75 mg/100 ml) has a mol. wt. 1,000,000 and is thus confined to the circulation or its local site of secretion. It has a half-life of about five days and γM antibodies have been proved particularly effective at destroying cell surfaces. Complement is essential to the final lysis and is fixed by the tails of the immunoglobulin while the arms are bound to antigen. With six units all in one place the activation of complement is probably more effective.

γD has a mean serum level of 30 mg/100 ml but little is yet known of its antibody activity.

ANTIBODY DEFICIENCY SYNDROME

Absence of γ-globulin by simple serum electrophoresis associated with recurrent infections which were prevented with γ-globulin therapy was first reported by Bruton in 1952. More sensitive methods invariably detect some γ-globulin and in a few cases the levels appear normal so that the term Antibody Deficiency Syndrome is best used for this first group of disturbances of the immunoglobulins.

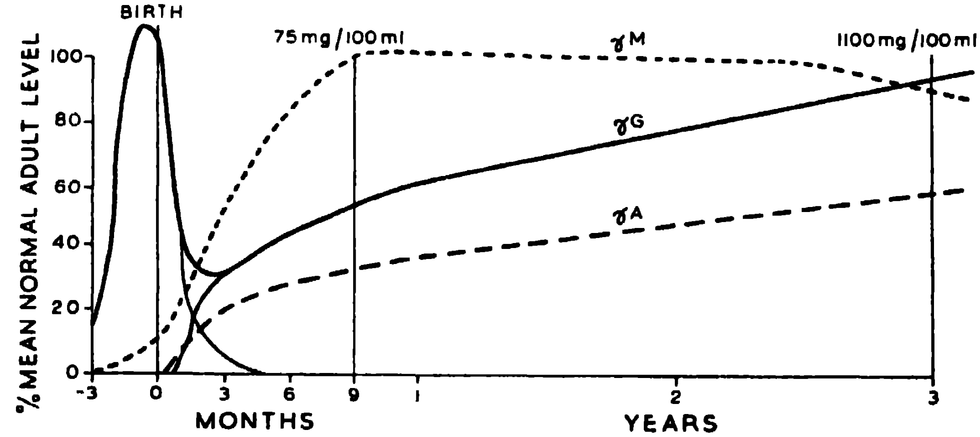


FIG. 3. The natural development of serum immunoglobulin levels. (After West *et al.*, 1962.)

Fig. 3 shows the natural development of the serum immunoglobulin levels. γG is transferred across the placenta mainly during the last trimester of pregnancy. This transfer can occur against the gradient, for we have seen mothers with levels of 120 and 200 mg/100 ml who at full term gave birth to babies with levels of 400 and 640 respectively. Some babies get off to a bad start by being born too soon and we can state that prematurity of ten weeks or more is likely to be complicated by recurrent infection. Dr Davies is at present assessing the prophylactic value of γ-globulin therapy in the premature baby unit at Hammersmith.

After normal birth the mother's contribution falls exponentially and the baby normally begins maturing at about 4–5 weeks to reach adult levels by about three years of age. Some infants are slow starters and may not mature for five years. During this take-over period low levels are so common that during the first year of life the Medical Research Council have defined their limit for hypo-γ-globulinaemia as below 100 mg/100 ml. Over one year the definition is below 200 mg/100 ml.

γM production usually matures much more rapidly—within nine months. There is usually some present at birth. Soothill (1962*b*) has shown this can be a useful pointer to prognosis—with a true genetic defect the γM usually remains very low, whereas with delayed immunological maturity, the slow starters, the γM levels are near normal. Nevertheless such slow starters suffering from infections can still qualify for tiding over with therapy until their rising γG levels show that they can look after themselves.

γA production is similarly affected, but its normal maturation is so slow, not gaining adult levels till adolescence, that it is not quite so useful a pointer.

This type of maturation is not seen in the germ-free animal, so that the environmental antigenic challenge must be an important factor. This may explain why West Africans in Nigeria have higher serum γ-globulin levels than we do in England, for when they have been in England for a year or more they have the same levels as us.

Table 1 attempts to give some idea of the incidence of the antibody deficiency syndrome at a general hospital. Some 6,000 patients have had their sera electrophoresed. Whenever a low γ-globulin was seen it was checked immunologically. 56 patients conformed to the Medical Research Council definitions of hypo-γ-globulinaemia, a further 82 were in a borderline range of 200–400 mg/100 ml. The 2 S.D. lower limit of normal is 600 and patients are described with symptoms referable to the lower range, 200–400. In 38 of the former and 35 of the latter infection was a real problem, which makes this syndrome the commonest of the disturbances of the immunoglobulins. Sub-group A, prematurity and delayed maturity, accounts for one quarter of our cases. In contrast we have a very limited experience of the hereditary types in sub-group B. These can occur with or without lymphoid hypoplasia, that is with or without a defect in cellular as well as in humoral immunity. The sex-linked recessive occurring in males has a minimum incidence of 1/100,000 and other types are rarer, so that an average

TABLE 1. Distribution of measured (totals) and symptomatic (*) cases detected by electrophoresis of sera from 6,000 new patients at Hammersmith Hospital. Referrals from other hospitals are not included

%*		ANTIBODY DEFICIENCY SYNDROMES	γG mg/100 ml			
			<200		200–400	
			Totals	*	Totals	*
	I	HYPOGAMMA-GLOBULINAEMIA				
24		A Physiological				
		Prematurity	2	2	10	8
		Delayed maturity	1	1	7	7
4		B Hereditary				
		Sex-linked recessive	1	1	—	—
		Non sex-linked	1	1	—	—
		Ataxia telangiectasia	—	—	1	1
		Others	—	—	1	—
16		C Secondary to protein deficiency				
		Malnutrition	—	—	2	2
		Malabsorption	2	2	2	1
		Nephrosis	4	1	16	2
		Renal failure	1	1	8	3
40		D Associated with R.E.S. diseases				
		Chronic lymphatic leukaemia	7	7	2	—
		Lymphoma	1	1	—	—
		Prelymphoma	1	—	—	—
		Macroglobulinaemia	2	1	2	—
		Myeloma	17	9	4	1
		Premyeloma	2	2	—	—
		Hodgkin's	2	2	2	1
		Thymoma	1	1	—	—
		Prethymoma	—	—	—	—
		Marrow aplasia	6	4	6	1
		Extensive marrow disease, ca. etc.	3	1	5	0
		E Idiopathic Familial predisposition Truly acquired	2	1	14	8
			56	38*	82	35*
1	II	IATROGENIC Excess steroids, X-ray or cytotoxics	1	1	—	—
	III	WITH APPARENTLY NORMAL IMMUNOGLOBULINS				

hospital is only likely to see an affected family once in ten years. It is only with nation-wide surveys, such as the current Medical Research Council trial, that the full value of these 'experiments of nature' can be assessed, Squire (1962) and Good, Kelly, Rotstein and Varco (1962).

In sub-group C the mechanism seems simple, however in the nephrotic syndrome the actual γG lost in the urine only occasionally is adequate explanation of the low serum level. More often it is less than half the normal daily output of around 2·5 g, so that there must also be impaired synthesis of γG, probably secondary to the shortage of building materials. With malabsorption the relationship can be more complicated (Swift, 1962). We have good evidence that impaired γG synthesis is an important factor underlying the proneness to infection seen in uraemia.

D is the largest group, even though I have only included patients presenting from within the hospital—for we do attract three times as many more from outside. Apart from obvious diseases of the reticuloendothelial system, such as chronic lymphatic leukaemia, lymphoma, macroglobulinaemia, myeloma, Hodgkin's, thymoma and marrow disorders, we have seen three patients where the hypo-γ-globulinaemia was shown one or two years before the reticulosis appeared clinically. Dancaster, Hussain and Jackson (1959) described this as premyelomatosis in a patient with a long history of recurrent infection before she developed clear-cut myelomatosis. Prelymphoma has been similarly described, and Wollheim, Belfrage, Coster and Lindholmn (1964) recently reported a case observed for two years before any evidence of thymoma could be detected. Which really does come first? A neoplastic change which inhibits normal immunoglobulin production, or a failure of humoral immunity which results in overstimulation of the cellular elements to a degree where hyperplasia fades into neoplasia. I expect, as usually, it will be a bit of both.

Good evidence of genetic factors now exists in many cases of hypo-γ-globulinaemia acquired out of the blue late in life (Wollheim *et al.*, 1964). Such factors could be analogous to the late failure of intrinsic factor seen in pernicious anaemia, indeed autoimmune phenomena are frequently seen in these patients, but it is a little hard to conceive of plasma cells committing suicide. Many cases show an isolated absence of one or other of the immunoglobulins and this is also found in their families. Ataxia telangiectasia represents a half-way house for these children are born with inability to make γA, but γG production matures normally and only begins to fail around five years of age. However their relatives and other people have had isolated γA or γM deficiency and yet have lived a normal life, so familial predisposition can't be the whole story and I have no doubt that much hypo-γ-globulinaemia is truly acquired.

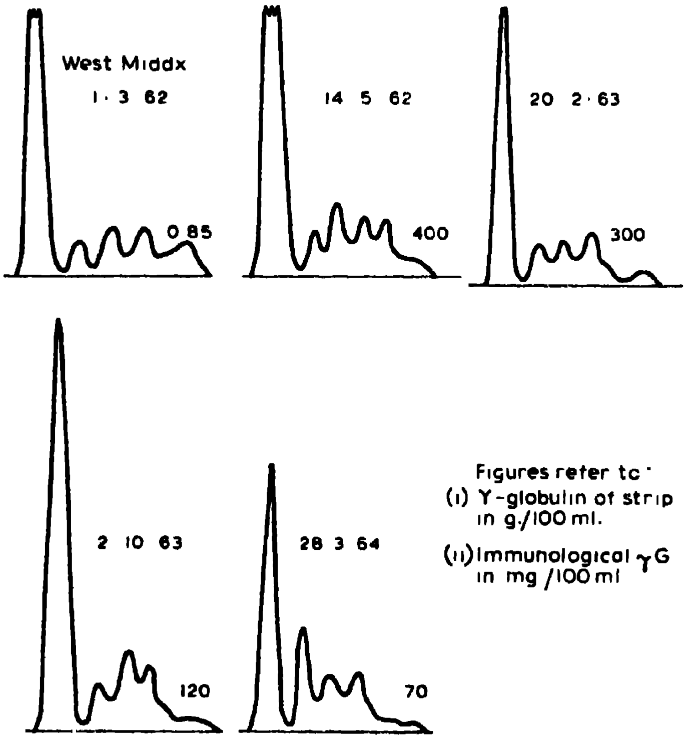


FIG. 4. Acquisition of hypo-γ-globulinaemia.

Fig. 4 shows the actual observation of such acquisition, and all the immunoglobulins were initially within normal limits. Thanks to Professor Dacie who has kept aliquots of serum deep frozen for many years, we have now got some six similar pedigrees. Another Hammersmith patient who developed a thymoma at 70 years of age, had a twin proven to be identical. This twin had normal immunoglobulins, thus excluding genetic factors.

Plate III

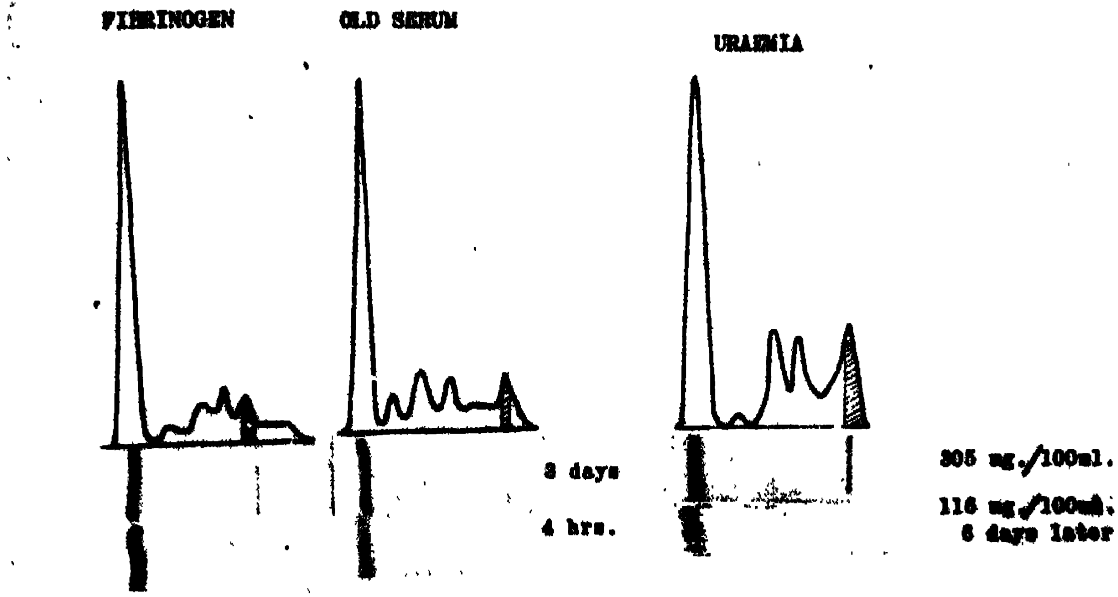


Fig. 5. False M bands.

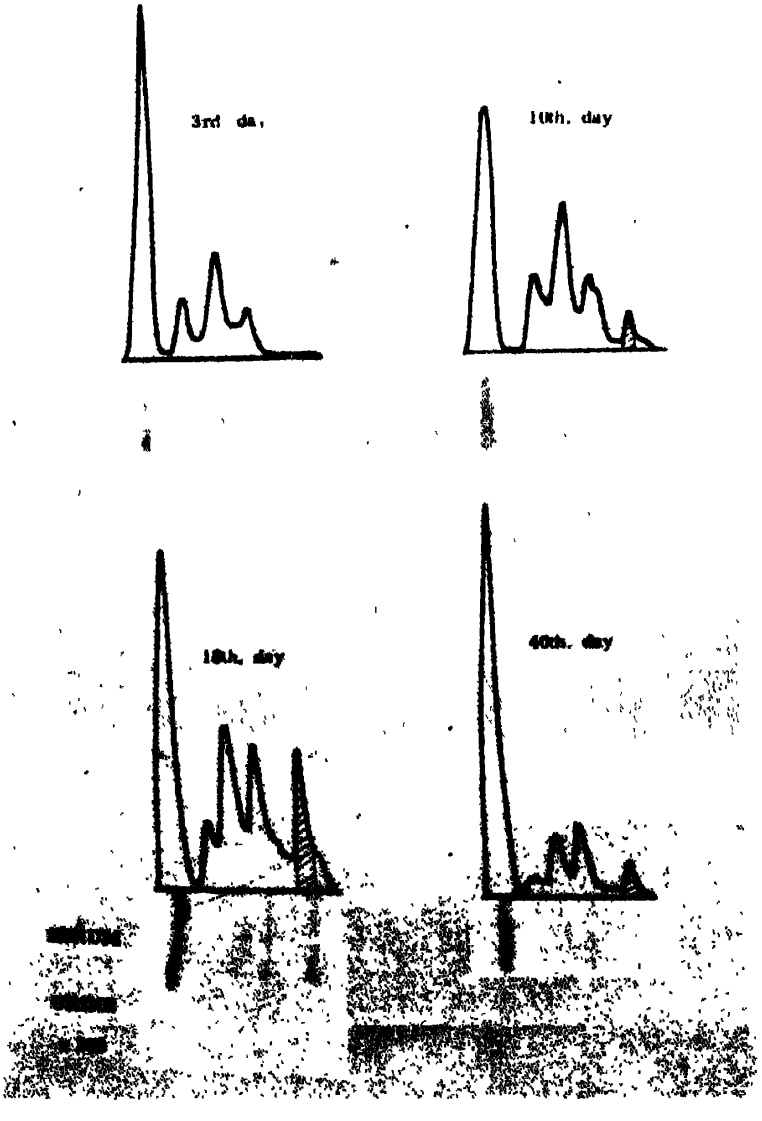


Fig. 6. A transient M band, readily seen against the pre-existing hypo-γ-globulinaemia in the sera from a patient with chronic lymphatic leukemia who developed pneumococcal pneumonia. The pneumonia and M band resolved in parallel.

PLATE IV

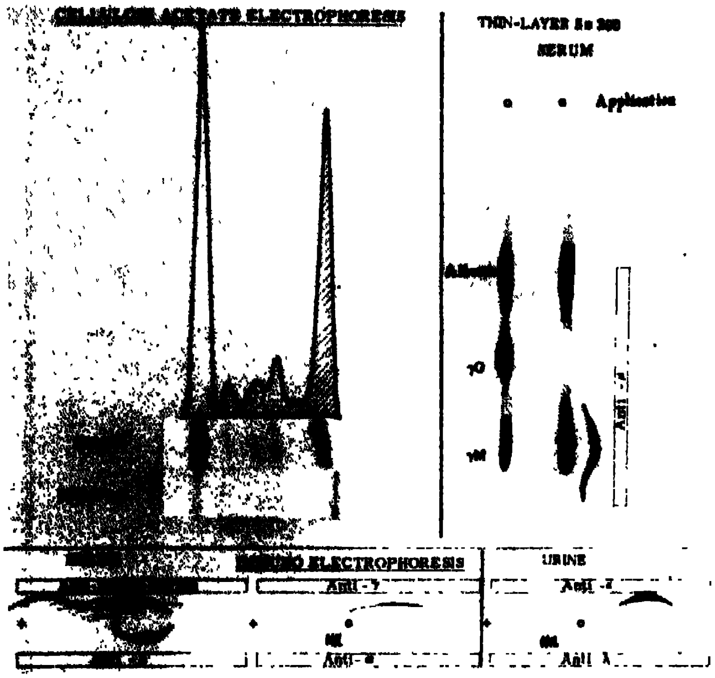


FIG. 7. Three methods used routinely to investigate M bands, (*i*) By running suitably concentrated urine alongside the serum, most Bence Jones proteins are detected as bands of a mobility different from that of the serum M. (*ii*) Immunoelectrophoresis of the urine confirms the B chains as of one type only. The serum M band is also identified by sera monospecific to the Λ chains. (*iii*) Thin-layer Sephadex G200 enables rapid confirmation of the main molecular size as comparable to γM, γG or smaller, and can be combined with monospecific antiserum.

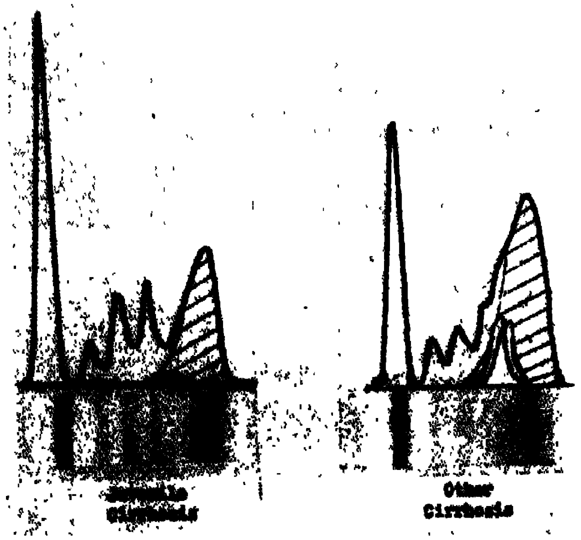


FIG. 8. Polyclonal gammopathy, on the left, shows absense of β-γ fusion because proteins with this mobility are normal (γA 280*, γM 80*) or reduced (β_1 C, complement) and the increase is mainly in slow γG (2,900*). In contrast healthy reactive increases show β-γ fusion, involving γA (960*), γM (240*) and γG (3,100*) with no reduction of β_1 C. (*Means of results in mg./100 ml. from 40 patients roughly paired by the height of the γ peak.)

In group II patients given excess steroids, whole body irradiation or cytotoxic drugs, subsequent infection often ruins the object of the treatment. We have looked many times but have only proven hypo-γ-globulinaemia once, so other factors are involved.

Group III is important because it can be readily missed, for on electrophoresis the serum pattern looks normal. On measurement the immunoglobulin levels are within normal limits, although again there is a half-way house and many patients have shown isolated deficiency of γM or γA. Yet there is a clear history of repeated infection. A patient of Dr. Citron had had some 200 chest infections, yet when given prophylactic γ-globulin treatment these were markedly reduced. Such is the convincing evidence, that although his own immunoglobulins may have looked normal (γG^{640} γA^{240} γM^{75} mg/100 ml) they didn't work. Very little antibody could be assayed by standard methods. By challenging suspected patients with standard antigens this deficient antibody response can be unmasked. There would appear to be something wrong with the manufacture of the business end of the immunoglobulin molecules which carries the antigen combining sites.

Thus in patients with the antibody deficiency syndrome, failure is not complete, there is an *immune paresis* rather than a paralysis. In most cases the defect is visible as hypo-γ-globulinaemia, and simple assay of the immunoglobulins is diagnostic. In a few cases the immunoglobulin levels appear adequate and the defect can only be detected by testing the ability to produce antibody.

MONOCLONAL GAMMOPATHIES

We recognize the second main group of immunoglobulin disturbances by the appearance of a narrow monoclonal band in the electrophoretic strip. Riva (1957) called them M bands to cover their occurrence in myelomatosis, macroglobulinaemia and malignant diseases, and Waldenstrom (1962) suggested we keep this term for monoclonal.

Cohen and Porter (1964) have produced important evidence that M bands are largely monoclonal. If the isolated immunoglobulin is reduced and electrophoresed under conditions unfavourable to polymerization, its A and B chains can be separated. Normal γG produces a whole family of B chains, which at neutral pH appear as at least ten bands. In contrast M proteins of all types produce a homogeneous B chain, a single band or at the most two. Antigenically this is always only one type, whereas the normal mixture contains both K and L. One of the antibodies isolated from guinea-pigs given hapten antigens by Edelman's group (Edelman, Benaceraff and Ovary, 1963) produced an identical picture which further supports a monoclonal hypothesis. In myelomatosis, the Bence-Jones proteins match the B chains of the M protein and there is now good evidence that they are produced by the same cells that produce the parent protein. They reflect an inefficient system making more B than A chains.

Checking M bands by immunoelectrophoresis avoids being misled, see Plate III, Fig. 5.[1] Sometimes blood specimens clot badly and some fibrinogen is left behind, with care it can be eliminated. Samples left over the weekend, or even overnight, often show a narrow zone or two in the middle or at the end of the γ-globulin. This is probably the result of dimer formation and subsequent breakdown (James, Henney and Stanworth, 1964). It is not often seen in fresh serum except from patients with uraemia, where with remission, it disappears.

When it was claimed (Hallen, 1963) that 1 per cent of the general population and 3 per cent over 70 years had M bands in their serum I was surprised. However, a survey of 396 patients over 70 years did yield the 3 per cent. Three were from myeloma cases, but nine seem to be pursuing the benign course emphasized by Waldenstrom (1964) when he described these as essential monoclonal gammopathies. These M bands though minimal have all been confirmed by a typical M bow on immunoelectrophoresis and all three immunoglobulin types have been found, however the *essential* features are:

(1) on the whole they are present in low concentration, which over the years increases very slowly if at all.

(2) there is no loss of background γ-globulins or other evidence of immune paresis.

(3) Bence-Jones counterparts are not found in the urine, even when it has been concentrated over 300 times.

---

[1] The plates referred to in this paper will be found between pages 114–15.

I think they are seen better on cellulose acetate and our paper results may explain why we have been missing them in the past.

My interest in M bands was fired in Professor Maclagan's department at the Westminster Hospital and Brian Cooke and I have been collecting them for some seven years (see Table 2).

TABLE 2. Diagnosis achieved in 456 patients with M bands in their sera

1. MALIGNANT PROLIFERATION OF R.E.S.		85%
Myeloma	67%	
Macroglobulinaemia	12%	
Leukaemia, lymphoma, etc.	6%	
2. BENIGN (ESSENTIAL) Long observation		6%
3. CANCER		4%
4. OTHER DISEASES Liver, gut, etc.		4%
5. TRANSIENT		1%

Series of Waldenstrom *et al.*, 1961 (216) with Cooke and Hobbs, unpublished (240)

Diagnosis not achieved in 87 further patients, mainly due to short observation

Combining our results with Waldenstrom's (1961) irons out some differences. In many cases no explanation has been found, usually because it takes 5–10 years for the underlying process to be recognized as a clinical syndrome. Where a diagnosis was achieved, the commonest lesion was malignant proliferation within the RES. Mostly this is recognized as myelomatosis, which remains a clinical diagnosis with involvement of bones by definition, but as with biological definitions nothing is watertight and these shade via the soft tissue plasmacytoma, through macroglobulinaemia to frank neoplasia in the lymphoid tissues even spreading into the blood stream to be classified as leukaemia. What I believe is that neoplastic mutation of some stem cell has led to a single clone of cells capable of producing immunoglobulin but only the cells doing so take on the appearance of plasma cells. Generally such clones behave in a malignant manner and invade other tissues, but in some cases invasion is not seen over 20 years, none of these clinical sequelae occur and such cases are classified as essential. No doubt in the future their percentage is going to be much increased.

M bands are also found where a cancer in other tissues or other diseases, such as cirrhosis, dominate the clinical picture. Lesions of the gut seem especially associated, indeed most of these cancers were of the gut. The gut is naturally rich in lymphoid tissue and Osserman, Takatsuki and Talal (1964) believe the association of gut neoplasia with monoclonal 'stem cell' neoplasia could be the result of mutual exposure to a common carcinogenic factor. Waldenstrom (1964) has suggested the M band might be antibody against the cancer but this has never been shown. Monoclonal antibody better explains the rare transient variety.

A patient with chronic lymphatic leukaemia had had the usual hypo-γ-globulinaemia over nine months despite two previous infections. Then he was admitted with pneumococcal pneumonia and on the third day of his illness his strip showed no change, see Plate III, Fig. 6. His progress was stormy despite antibiotics; by the tenth day he had produced a small M band, which reached a peak at 18 days then declined to a low level at 40 days and has now almost disappeared. The pneumonia slowly resolved in parallel. This M band was γG and only had B chains of type L. The timing of this whole picture suggests a single clone capable of responding to one of the many pneumococcal antigens survived the immune paresis in this patient. The low background γ enabled us to watch it respond—and efficiently, for there were no dud B chains in the urine. Alas we have not yet identified the antigen to prove the antibody.

TABLE 3. Immunological type Material from 386 patients with undoubted myelomatosis

γG	67%
γA	25%
γμ (BENCE-JONES)	6%
γM	1%
No paraprotein detected	1%

Two M bands in 2%, Cryoglobulins in 2%, γμ (heavy chain) <0·5%
Series: Waldenstrom *et al.*, 1961 (129), Drivsholm, 1964 (105), Cooke and Hobbs, unpublished (162)

Many series of patients with myelomatosis have been reported, but only recently has this disease been studied in the light of its immunological type, see Table 3. 67 per cent are

γG, 25 per cent are γA and 1 per cent are γM with bony lesions. While these neoplastic clones can produce immunoglobulins comparable to normal proteins, and thus called *paraproteins*, they do so inefficiently in some 65 per cent of cases as judged by the associated Bence-Jones proteinuria. In 6 per cent of cases they make such a hash of it that all we can find are B chains. In a rare type of dyscrasia B chains are not detectable and only fragments of A chain are found—four such patients have been reported including Franklin's original case (Franklin, Lowenstein, Bigelow and Meltzer, 1964). Finally we do sometimes see cases with extensive masses of neoplastic cells indistinguishable from plasma cells, yet where with the most refined techniques no M components can be detected in serum, urine or even in the tumour which must represent the last stage of dedifferentiation. This sort of typing does have some bearing on the disease pattern (Drivsholm, 1964). γG myelomata are those most frequently complicated by recurrent infection as a result of associated immune paresis. γA myelomata in contrast are more often complicated by hypercalcaemia and renal failure. $\gamma\mu$ myelomata also result in renal failure. γM paraproteinaemia underlies most cases of the viscosity syndrome.

TABLE 4. The clinical effects associated with paraprotein production

1. BONE LESIONS	Lack of osteoblastic reaction
2. ANAEMIA	Normocytic normochromic (? haemodilution)
	Leucoerythroblastic
	Megaloblastic, B12 or Folate deficiency
3. IMMUNE PARESIS	(Frequent and severe with γG type)
4. RENAL FAILURE	Pyelonephritis
	Hypercalcaemia
	Myeloma-kidney
	Amyloid nephrosis
	No visible lesion
	Amino-aciduria, tubular acidosis, invasion
5. Amyloidosis	
6. Cryoglobulinaemia (see MacKay *et al.*, 1956)	
7. Viscosity syndrome, purpura, coma	
8. Hyperlipaemia, xanthomatosis (see Lennard-Jones, 1960)	

Table 4 shows the effects associated with paraprotein production. For some unknown reason myeloma invasion of bone nearly always shows a lack of osteoblastic reaction, so that a normal serum level of alkaline phosphatase is the rule, and a useful diagnostic pointer to the disease. The commonest type of anaemia is normocytic and normochromic, and where this occurs unexpectedly in the aged the diagnosis of myeloma should always be considered. It has been suggested that haemodilution, subsequent to the osmotic effect of the paraproteinaemia, may be an important factor. Leucoerythroblastosis is seen late when the normal marrow has been displaced by extensive myeloma deposits. Megaloblastic anaemia is also associated with myelomatosis and may be mediated by either B12 or folate deficiency. The latter seems commoner and may result from excessive cell growth, for it is often seen with Bence-Jones proteinuria which can readily increase to ten times above the normal daily production of immunoglobulins.

It has been suggested that immune paresis results because the normal plasma cells cannot compete with the misappropriation of amino acids by the tumour cells, for there is some correlation between the height of the M band and the whiteness of the background. However, typing reveals it is most frequent and most severe in patients with γG myelomata, which only achieve the highest serum levels because γG has the longest half-life. The γG production rarely exceeds 20 g/day whereas we have seen several Bence-Jones myelomata producing up to 50 g/day, with only moderate reduction of serum immunoglobulins.

Another mechanism could be marrow displacement, but leucoerythroblastic anaemia is only seen late in the disease, and we have seen hypo-γ-globulinaemia precede detectable tumour in two patients. Wieme has one patient who presented with pneumonia; the chest X-ray revealed a walnut-sized cavity in the sternum but no other bone lesions were detected. Wieme found hypo-γ-globulinaemia and detected traces of M protein. The tumour was excised and confirmed as a myeloma. Three months later the patient's γ-globulin returned to normal and three years later she was still alive and well. We have seen one similar case where the remission has lasted two years following the removal of a solitary myeloma from the head of the femur. These two cases suggest that the tumours released some humoral factor whereby the normal plasma cells were suppressed.

Nowadays renal failure is the mode of death of about half the patients with myeloma. Pyelonephritis, aggravated by immune paresis is the commonest lesion closely followed by 'myeloma-kidney', where the distal tubules are characteristically plugged by casts which frequently stain like amyloid. The casts most commonly represent complexes of Tamm-Horsfal with Bence-Jones proteins, but γA myelomata also frequently show an excessive urinary loss of γA and I have found it in casts. Thus myelomata with associated proteinuria carry a poor prognosis and it is also known that myeloma-kidney can be precipitated by intravenous pyelography, the concentration of the radio-opaque substance encourages cast formation. Other renal lesions result from hypercalcaemia (especially if the patients become bed-ridden—to be avoided if possible), amyloidosis and very rarely invasion of the kidney by myeloma cells. In some cases the patient dies of renal failure and post-mortem histology reveals no obvious lesion.

Amyloidosis occurs in some 5–10 per cent of myeloma patients and characteristic sites are around the collagen of the carpal tunnel and the tongue. Renal involvement usually results in a nephrotic syndrome. Whenever unexplained amyloid occurs it is imperative that the urine be examined, even if it has to be concentrated some 300 times, the more so when the serum shows no M band or immune paresis only. Thin-layer Sephadex, G 100 (*Morris*, 1964) can help to identify the monomers (mol. wt. 22,000) and the dimers (mol. wt. 44,000) of Bence-Jones proteins compared with a haemoglobin marker (mol. wt. 65,000) and confirmed by specific antisera against K or L chains. Osserman *et al.* (1964) have detected monoclonal B chains in seven cases of non-myeloma amyloid and they would seem to be the common factor in all types of amyloid. Only a few Bence-Jones proteins are associated with amyloid, and Osserman and his co-workers have shown that these have a higher affinity for human tissues than do B chains from non-amyloid sources. It is believed that the incorporation of such atypical B chains into the tissues is fundamental in the formation of amyloid. Such atypical B chains are the product of abnormal plasma cells, resulting from a mutation which could be primary as in myeloma or idiopathic cases, genetic as in familial varieties or secondary to the prolonged antigenic overstimulation of chronic disease.

In the viscosity syndrome, the chief complaint is lassitude. The fundus oculi typically shows sausage-shaped distension of the retinal veins, thought to result from sludging of the viscous blood, and in severe cases perivenous haemorrhages can lead to blindness. Examination of the blood typically reveals anaemia and an E.S.R. above 100 mm/hr. A water dilution test of the serum is usually positive, and on electrophoresis the M band is usually scalloped (see Plate IV, Fig. 7) though this is not a reliable feature. The excess of paraprotein coats the platelets, impairing their function so that purpura is common; the red cells, accelerating their destruction in a non-specific manner; and the white cells, interfering with phagocytosis to aggravate any immune paresis (Pachter, Johnson, Neblett and Truant, 1959). While most cases result from γM production, γG and even one Bence-Jones protein have polymerized to a degree resulting in the same syndrome (Waldenstrom, 1961). Thus immunoelectrophoresis most often shows a γM bow (see Plate IV, Fig. 7). The typical Waldenstrom type of macroglobulinaemia involves the marrow diffusely (only rarely with true punched out myeloma deposits), the spleen moderately and the lymph nodes only to a minor degree. In such cases the precipitin arc curves back in a normal manner, see Plate IV, Fig. 7. Mackay, Taft and Woods (1957) described macroglobulinaemia as a secondary phenomenon, possibly a further mutation, occurring in pre-existing lymphosarcomata where the disease mainly involves massive lymph-nodes. In such cases Clausen (1963) has noted that the precipitin arc curls back to the antiserum trough in a most unusual manner and we have confirmed this in seven cases. Another useful inexpensive method uses thin-layer Sephadex G 200 instead of the ultracentrifuge (see Plate IV, Fig. 7). The largest molecules do not enter the dextrangel and run furthest in the free solvent. Smaller molecules enter the gel to a varying extent and are slowed in order of size, the smallest being closest to the origin. After the run, a dry strip of cellulose acetate is applied and the pattern becomes transferred. The strip can then be removed and stained in Ponceau-S to show the proteins. The known γG myeloma on the left shows its peak in the middle; the γM, its peak at the bottom. By laying a second dry cellulose acetate strip on to the first before staining the pattern is transferred into that strip also. A thin strip of filter paper soaked in antiserum can then be applied to the second strip and the immune precipitate can be allowed to develop under liquid paraffin for three days. The residual proteins can then be washed out and the precipitin arc stained with Ponceau-S. The original strip is stained as a master. This method thus combines molecular sizing with monospecific antiserum and gets the best of both techniques within the capabilities of most laboratories.

POLYCLONAL GAMMOPATHIES

The third group of immunoglobulin disturbances are the polyclonal gammopathies. Their definition is really a confession of ignorance—a strip γ-globulin raised above 1·3 g/100 ml without any known exogenous antigenic stimulation.

In Laennec's cirrhosis and indeed in most cases of chronic infection such as bacterial endocarditis, tuberculosis or kala-azar the increased immunoglobulin is truly polyclonal; a broad spectrum involving γG, γA and γM which are all greatly increased. This results in an electrophoretic picture with β-γ fusion, this area being filled up by γA and γM. All these increases are a healthy reaction to the associated disease with its many exogenous antigens. In cirrhosis the rise is inversely proportional to the albumin level, and can be loosely regarded as a compensation—a healthy attempt by the reticulo-endothelial system to maintain the serum protein level. In any event all three immunoglobulins are involved and the picture is truly polyclonal, see Plate IV, Fig. 8. In contrast the increases found in the so-called polyclonal gammopathies fit into this picture. They mainly involve γG only and γA and γM show normal levels. In juvenile cirrhosis the γG can still be grossly elevated when the albumin is normal. The other characteristic is the anti-complementary activity of the serum, so that β_1C is deficient. The sum effect of this is an absence of β-γ fusion in the strip. In old sera β_1C changes its mobility and the β_2 goes down so it is important to run the serum within four hours of collection, to know that a reduced β_2 is genuine. On cellulose acetate the distinction is easily made visually—look at the pale gap, or where the γ fuses right into the β. It gave a 95 per cent reliable separation of juvenile from other cirrhoses in 200 cases.

TABLE 5. Polyclonal gammopathies

Symptomatic
Juvenile cirrhosis (chronic active hepatitis)
Hyperglobulinaemic purpura
Disseminated lupus erythematosis
Chronic sialoadenitis
SOME Hashimoto's thyroiditis
Rheumatoid arthritis
Reticuloses, lymphosarcoma, sarcoidosis, etc.
Collagenoses, polymyositis, scleroderma, etc.
Henoch-Schönlein syndrome, etc.
Essential
Often familial

Table 5 shows the clinical diseases which have been included in the polyclonal gammopathies—juvenile cirrhosis, hyperglobulinaemic purpura, D.L.E., chronic sialoadenitis and cases of various diseases when they show a markedly raised γ-globulin. In nearly all our cases like this where the immunoglobulins have been measured, the polyclonal increase has been only in γG, and the strip has shown absence of β-γ fusion. The other common features are:

(i) we don't understand them,
(ii) they often do well on steroids,
(iii) many are thought to have an autoimmune basis.

Autoimmune reaction implies a breakdown of our normal tolerance of our own tissue antigens. Burnet believes that tolerance is achieved by the elimination of potentially reactive cells before immunological maturity occurs. Any subsequent reactivity would only arise through new clones, or new antigens. If a wild clone arose, surely we would expect its antibody to be monoclonal. On the other hand, an alteration in antigenicity could be expected to produce a narrow spectrum of antibodies which might initially be of only one type of immunoglobulin as in this limited polyclonal picture.

Leonhardt and Larsson (1959) undertook a large survey of some families in which several cases of so rare a disease as

D.L.E. had occurred and showed statistically that such families had higher γ-globulin levels and more D.L.E. than the general population. Most of the relatives were asymptomatic, essential, but after periods of ten years or so became symptomatic—sisters both developed D.L.E. Members of these families may be ultrasensitive in detecting minor changes in tissue antigenicity, and may thus be readily triggered into autoimmune reactions.

In similar families, other workers have found an increased incidence of hypo-γ-globulinaemia. If a common genetic defect is to account for too little or too much immunoglobulin it will have to be in something like an operator gene.

SUMMARY

Immunoglobulin disturbances are recognized in three main groups, deficiencies and monoclonal and polyclonal excesses.

(A) Primary antibody deficiency syndromes are rare and may arise as:

1. A failure of B chain synthesis resulting in a deficiency of all immunoglobulins.
2. A failure of A chain synthesis resulting in isolated deficiencies, often symptomless.
3. A failure to make effective antigen combining sites.
4. Permutations of these.

Secondary antibody deficiency syndromes are common, and may be treated more often as supplies of γ-globulin improve. Humoral agents from reticuloendothelial neoplasms may be important causative factors.

(B) There is strong evidence of a monoclonal nature in some gammopathies, notably myelomatosis where γG clones tend to be associated with severe immune paresis, γA clones with hypercalcaemia and renal failure and γM clones with the viscosity syndrome. However there is overlap among these patterns. Atypical B chains seem to be important in the pathogenesis of amyloidosis.

(C) The polyclonal gammopathies are rather nebulous, but can have an hereditary basis and in contrast to the broad spectrum of healthy responses are mainly γG only with absence of β-γ fusion. This narrower spectrum may reflect a response to autoantigens.

Finally most of the new ideas in this ever-widening field would not have been possible had it not been for the recent fundamental work on immunoglobulin structure.

REFERENCES

Bruton, O. C. (1952) *Pediatrics*, **9**, 722.

Clausen, J. (1963) *Science Tools*, **10**, 29.

Cohen, S. and Porter, R. R. (1964) *Adv. Immunol.*, **4**, 287.

Dancaster, C. P., Hussain, O. A. and Jackson, W. P. U. (1959) *Postgrad. med. J.*, **35**, 662.

Drivsholm, A. (1964) *Acta med. scand.*, **176**, 509.

Edelman, G. M., Benaceraff, B. and Ovary, Z. (1963) *J. exp. Med.*, **118**, 229.

Franklin, E. C., Lowenstein, J., Bigelow, B. and Meltzer, M. (1964) *Amer. J. Med.*, **37**, 332.

Good, R. A., Kelly, W. D., Rotstein, J. and Varco, R. L. (1962) *Progress in Allergy*, **6**, 187. Karger: N.Y.

Hallen, J. (1963) *Acta med. scand.*, **173**, 737.

Heremans, J. F. (1959) *Clin. chim. Acta*, **4**, 639.

James, K., Henney, C. S. and Stanworth, D. R. (1964) *Nature, Lond.*, **202**, 563.

Lennard-Jones, J. E. (1960) *Brit. med. J.*, **i**, 781.

Leonhardt, T. and Larsson, O. (1959) *Acta med. scand.*, **165**, 371.

Mackay, I. R., Eriksen, N., Motulsky, A. G. and Volwiler, W. (1956) *Amer. J. Med.*, **20**, 564.

Mackay, I. R., Taft, L. I. and Woods, E. F. (1957) *Brit. med. J.*, **i**, 561.

Morris, C. J. O. R. (1964) *J. Chromatogr.*, **16**, 167.

Osserman, E. F., Takatsuki, K. and Talal, N. (1964) *Seminars in Haematology*, ed. Miescher, P. A., p. 3. Grune & Stratton: N.Y.

Pachter, M. R., Johnson, S. A., Neblett, T. R. and Truant, J. R. (1959) *Amer. J. clin. Path.*, **31**, 467.

Riva, G. (1957) *Das Serumeiweissbild.* Verlag. Han. Huber, Bern.

Rogers-Brambell, F. W. R., Hemmings, W. A. and Morris, I. G. (1964) *Nature*, **203**, 1352.

Rowe, D. S. and Fahey, J. L. (1965) *J. exp. Med.*, **121**, 171.

Scheiddeger, J. J. (1955) *Int. Arch. Allergy*, **7**, 103.

Singer, S. J. (1965) *Immunochemistry*, **1**, 15.

Soothill, J. F. (1962*a*) *J. Lab. clin. Med.*, **59**, 859.

Soothill, J. F. (1962*b*) *Clin. Sci.*, **23**, 27.

Squire, J. R. (1962) *Proc. roy. Soc. Med.*, **55**, 393.

Swift, P. N. (1962) *Postgrad. med. J.*, **38**, 633.

WALDENSTROM, J. (1961) *Acta med. scand.* suppl., **367**, 110.
WALDENSTROM, J. (1962) *Progress in Allergy*, **6**, 320. Karger: N.Y.
WALDENSTROM, J. (1964) *Acta med. scand.*, **176**, 345.
WEST, C. D., HONG, R. and HOLLAND, N. H. (1962) *J. clin. Invest.*, **41**, 2054.
WIEME, J. Personal communication.
WOLLHEIM, F. A., BELFRAGE, S., COSTER, C. and LINDHOLMN, H. (1964) *Acta med. scand.*, **176**, 1.

VIII

Rheumatoid Factor

JOHN R. SQUIRE

WITH the great public interest in scientific medicine, one is sometimes asked by journalists for information about this or that 'break-through'; indeed, some of our friends would like to have the television cameras focused on just those initial experiments which will lead to considerable extensions of knowledge. It is difficult to see how to oblige, when a prime discovery like that of Waaler (1940) passes almost unnoticed. Even at the time when Rose, Ragan, Pearce and Lipman (1948) confirmed and extended the idea that a new serological test was available for the study of rheumatoid arthritis, few authorities would have spoken of a 'break-through'. In retrospect, it can be seen that investigations based on the apparently simple system of the Waaler-Rose test (or Rose-Waaler or Rose test—all synonyms) have increased our understanding not only of one particular group of diseases, but also of the behaviour and structure of several immunoglobulins; it is now even contributing to an important branch of genetics.

Excellent reviews are already available (Glynn, 1963; Rowe, 1964), yet progress is still so rapid that further evaluation of current work is justified. It is also worth re-emphasizing the importance of accidental observations about human disease, since in the hands of informed observers they often point to a trail of discovery. The rewards come, of course, from systematic tests which must then be made of all the relevant features; the length of the follow-up period—in this case, decades—is largely a function of the interdependence of the many different branches of medical knowledge—immunology and biochemistry, for instance, which progress simultaneously.

The Waaler-Rose reaction depends on a simple system, familiar for many years in connexion with complement-fixation reactions (indeed, Waaler was attempting to carry out such a test). Sheep red cells are injected into a rabbit to excite antibody formation. The antiserum can then be used to 'coat' sheep erythrocytes with antibody; when antiserum (previously subjected to mild heat to destroy complement) is used in suitable dilution, this coating process leads to no visible change in the erythrocytes, nor does any change follow the addition of normal human serum. But after adding serum from most patients with rheumatoid arthritis, suspensions of these antibody-coated erythrocytes are strongly agglutinated. Furthermore, many rheumatoid sera show agglutination in this system when diluted 1,000-fold or more, that is to say, a 'high titre' of 'rheumatoid factor' is present.

This sheep-cell agglutination test is still used substantially as described. To avoid errors due to 'heterophile antibody' (present in human sera which clump sheep erythrocytes, coated or not), it has been found useful, in investigational work at least, to absorb the human test sera with uncoated sheep red cells before performing the test. Appropriate controls with uncoated cells and with known positive and negative test sera are also always included. Other improvements which can now be suggested are noted later.

To follow developments in knowledge step by step in time sequence is of great interest to the specialist. For general purposes, it is more appropriate to give the results, with some indication of the procedures which have led up to them. The principal topics which have contributed to our understanding the nature of rheumatoid factor and its behaviour are:

(1) the recognition of different classes of immunoglobulin. Four are now known, and at least two of these are concerned in the sheep cell agglutination test.

(2) the knowledge of the polypeptide chains and of their fragments which can be derived from immunoglobulins by reduction and by proteolytic digestion.

(3) the concept of autoantibody formation and the possible significance of allotypes in the pathogenesis of this kind of process.

Classes of Immunoglobulin

In a recent lecture on immunoglobulins (Squire, 1964), it was said that too many names existed for each of the three classes then known. Now an internationally agreed terminology has been put forward (*Bull. Wld. Hlth. Org.*, 1964), and a fourth immunoglobulin has been recognized. In Table 1, these names are given, together with the previous names which will for a while remain more familiar. The old names were of course based on the methods used for recognizing and separating these proteins. Thus, the old name 'γ globulin' refers to a range of electrophoretic mobility—the slowest-moving group of proteins at pH 8·6. But when immunoelectrophoresis was developed, it was seen that γ-globulin was heterogeneous in the sense that the main 'γ-globulin' line of precipitate extended well forward into the β- and even into the α_2-region. With this technique also, other different individual proteins were shown up as arcs of precipitate with overlapping electrophoretic mobilities. One of these was identified as the 'macroglobulin' with a high sedimentation ('S') constant, and became known as β_2M, γ_1M, etc. Another, not previously recognized, was similarly termed β_2A (or γ_1A in other laboratories). Ultracentrifugal separations had already divided the immunoglobulins into the main 7S γ globulin (mol. wt. about 150,000) and the 19S or γ macroglobulin (mol. wt. of the order of 1,000,000). But with the discovery of other immunoglobulins, the 7S class in the ultracentrifuge was shown to be heterogeneous, further complicating the old terminology. Replacing this confusion, the new nomenclature offers consistent descriptions, such as 'immunoglobulin G', written for short as 'IgG'; the others then become 'IgM', 'IgA' and now (cf. Rowe and Fahey, 1965) 'IgD'.

TABLE 1. The Four Immunoglobulins ('Ig') recognized—October, 1964

New names:	Previous names:
IgG	γ, 7Sγ, 6·6Sγ, γ_2, γ_{ss}
IgM	γ_1M, β_2M, 19Sγ, γ-macroglobulin
IgA	γ_1A, β_2A
IgD	—

(The alternative nomenclature proposed uses γG, γM, γA and γD, but this system appears to retain a little of the old ambiguity.)

INTERACTION OF IGM AND IGG AS THE BASIS OF TESTS FOR RHEUMATOID FACTOR

In this account, we are concerned only with IgM and IgG. As justified below, we may state that rheumatoid factor is usually in the IgM class, and is then as yet indistinguishable from any other human IgM except in terms of its (more or less) specific antibody-like reactivity. The rabbit antiserum used in the Waaler-Rose test also, of course, contains (rabbit) IgM and IgG, etc., and these can be separated in the ultracentrifuge or on columns of charged cellulose or 'Sephadex' gel. Most specimens of rabbit anti-sheep erythrocyte antiserum contain specific anti-sheep antibodies of at least two classes, IgG and IgM; either of these can be 'coated' on to sheep erythrocytes. *Only erythrocyte preparations coated with IgG can be agglutinated with rheumatoid factor.* Correspondingly, the rabbit antiserum usually contains at least two types of antibodies, capable respectively in full concentration of inducing haemolysis (with complement) or agglutination; the latter property is usually associated with the rabbit IgG, so that standardization of rabbit antiserum in terms of haemolytic power (primarily a function of the rabbit IgM) may give a false index of the potency of the antiserum for use in the Waaler-Rose reaction.

Since rabbit IgG is the essential component of the coated erythrocyte system for demonstrating rheumatoid-type IgM, it is evidently important to know whether IgG from other species would act in similar fashion. So far as animal sera are concerned, a number will serve, but with varying effectiveness. Horse IgG, for example, happens to be a rather poor indicator (this may be one reason why Waaler's observation was not made until 1940, since horse antisera are most commonly used to provide the 'amboceptor' in standard complement fixation tests). Of special interest, however, is human IgG and, in the period 1955–6, several new systems for the detection of rheumatoid factor were devised using human material. In some, the test depended on coating bacteria or red cells with IgG-antisera, e.g. anti-*Brucella* antibodies; incomplete Rhesus agglutinins. Others, more widely used, were based on purified IgG preparations from pooled plasma (e.g. Cohn fractions), which could be attached to tanned red cells or even simply adsorbed on to particles of clay (bentonite) or of plastic (polystyrene).

It might be expected that a simple system like plastic or latex rubber particles, which can be accurately sized and coated with purified human IgG preparations, would rapidly displace the somewhat more complicated Waaler-Rose reaction. This has not in fact happened, though in many laboratories, IgG-latex agglutination is used as a preliminary screening test. The reason for this depends on diagnostic specificity (Greenbury, 1960*a*). Even the Waaler-Rose test gives a small percentage of positive results with sera from apparently normal people; the IgG-latex test, however, gives many more 'false positives', though, admittedly, also giving positive reactions in a somewhat higher proportion of those diagnosed clinically as suffering from rheumatoid arthritis. The tests with human IgG preparations were also perhaps somewhat more variable according to the particular pool or Cohn fraction used. This difficulty prompted further tests with human IgG, the emphasis varying in different investigations between those designed principally to try to improve the *ad hoc* diagnostic test, and those seeking to understand how sera from a particular disease-group could react with a constituent (IgG) apparently already present in these sera.

Some batches of IgG preparations, it was found, actually form precipitates with rheumatoid sera, at least with those giving high-titre Waaler-Rose reactions. Stored preparations of IgG (usually about 90 per cent pure as made), show evidence in the ultracentrifuge of partial denaturation and polymerization, or of some fragmentation; in addition to the main 7S component, for instance, fractions can be found with 10S (higher molecular weight) or after longer storage, 3·5S characteristics (James, Stanworth and Henney, 1964). Edelman, Kunkel and Franklin (1958) tried the effect of deliberately damaging IgG preparations by the use of heat, alkali, etc. After this treatment, precipitate formation could be observed with all sera which contained rheumatoid factor. Initially it was considered that aggregation of the IgG up to some minimum size—certainly one of the effects of heat—was the essential prerequisite for precipitate formation with IgG alone, while in the other systems, the particles or erythrocytes coated by the IgG antisera provided the necessary steric arrangements for agglutination. This reasoning, however, still fails to explain—provided always that the rheumatoid IgM behaves like an antibody—why inhibition is not usually effected by the presence of native human IgG. But heat and other agents which cause denaturation lead to changes in the linkages between polypeptide chains; following Porter's original demonstration of the production of fragments of (IgG) gamma globulin with papain (1959), such remarkable progress has been made that quite reasonably precise suggestions can now be put forward, clarifying not only the nature of rheumatoid factor interactions, but also the relationships within the family of immunoglobulins.

STRUCTURE OF IMMUNOGLOBULINS

Various reducing agents (e.g. dimercaptoethanol) are effective in splitting disulphide (-S-S-) bonds, and when a purified immunoglobulin is treated in this way, separate polypeptide chains are obtained. Under certain circumstances, the removal of the reducing agent allows reaggregation of the chains to occur (not necessarily into the original pattern), but when the free sulphydryl (-SH) groups are blocked (e.g. with iodoacetamide), re-aggregation is prevented. The free polypeptide chains can then be separated in the ultracentrifuge into two major groups, the 'light' chains of molecular weight about 20,000 and the 'heavy', molecular weight about 50,000. In terms of concentration on a molar basis, approximately equal amounts of each group are obtained, so that an immunoglobulin of molecular weight of 150,000 (IgG, for example) is now regarded as being composed of two 'light' chains and two 'heavy' chains. A general structural formula for an immunoglobulin is shown in Fig. 1.

It has been known with certainty for three or four years now that antisera raised against one immunoglobulin of, say, human origin, will cross-react with preparations of another human immunoglobulin—in gel diffusion tests, for instance, 'reactions of partial identity' are obtained. These partial cross-reactions are due to the antisera containing a mixture of antibodies, some reacting with all immunoglobulins, some only with that used as antigen; by adsorption with the cross-reacting immunoglobulin, antisera can be obtained which are quite specific for a single immunoglobulin class. To take a particular example, most antisera raised by injecting an animal with human IgG, purified so as to be free from IgM, will nevertheless cross-react with IgM preparations which are free from IgG. But after the addition of suitable amounts of IgM to a sample of this antiserum and removal of the precipitate, this adsorbed antiserum will then only react with IgG and not with IgM. On the basis of inhibition tests with purified preparations of polypeptide chains, these antisera can be shown to be reactive only with the heavy chains derived from the particular immunoglobulin class with which they can form precipitates. To generalize on the basis of this and similar experiments, it can be stated that each immunoglobulin class is characterized by a particular type of heavy chain.

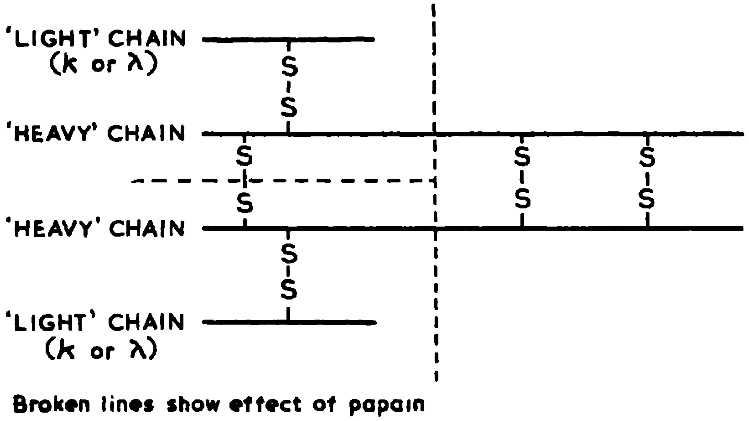


FIG. 1. All immunoglobulins appear to show a common 'atomic' structure. Two types exist in each class, depending on the nature of the 'light' polypeptide chain.

The light chain preparations, on the other hand, are indistinguishable, whatever class of immunoglobulin has served as their source. This statement remains exactly true for all immunoglobulins prepared from normal sera. In pathological states, however, such as myelomatosis or macroglobulinaemia, it is found that light chains derived from the abnormal protein separable from the patient's serum (or sometimes from the urine) can be used to produce antisera with a still narrower range of (light chain) reactivity. On this basis, using individual antisera raised against each of many such separated myeloma protein or macroglobulin antigens, normal immunoglobulins have been shown to fall into two main types, now to be known as type K or type L; both types are present in all normal immunoglobulin preparations. Normal IgG, then, is a mixture of type K IgG and of type L IgG, in contrast to a G-class myeloma protein from one individual patient which consists either of type K or of type L, but not of both. The corresponding light chains are now to be known as ϰ (kappa) or λ (lambda). All ϰ-chains (and λ-chains) from one normal human immunoglobulin class are identical (or perhaps closely similar) immunologically to the ϰ-chains (and λ-chains) derived from another class.

Reverting to the heavy chains, for consistency these are also to be designated by small Greek letters, and for the four immunoglobulins recognized to date, the γ-, μ-, α- and δ-chains represent the specific heavy chains corresponding to IgG, IgM, IgA and IgD classes of immunoglobulins (see Table 2). Thus, IgG can now be represented in terms of a molecular formula as $\gamma_2\varkappa_2$ for the K-type and $\gamma_2\lambda_2$ for the L-type. Normal IgG thus consists of a mixture of $\gamma_2\varkappa_2$ and of $\gamma_2\lambda_2$. So far as is known, the molecular weights of all heavy chains of whatever group are similar. It follows therefore that the higher molecular weight of

TABLE 2. Immunoglobulin polypeptide chains

(*a*)	Light chains	Ig type	Old name
	ϰ (kappa)	K	Type I
	λ (lambda)	L	Type II
(*b*)	Heavy chains	Ig class	Old name
	γ (gamma)	IgG	7Sγ, etc.
	μ (mu)	IgM	γ1M, etc.
	α (alpha)	IgA	γ1A, etc.
	δ (delta)	IgD	—

an immunoglobulin class such as IgM implies that there are considerably more polypeptide chains in this molecule. A reasonable formula for IgM is put forward as $(\mu_2\kappa_2)_6$ and $(\mu_2\lambda_2)_6$. The full list of molecular formulae is given in Table 3 and shown in structural terms in Fig. 2.

TABLE 3. 'Formulae' of immunoglobulins

	Type	
IgG	K	$\gamma_2\kappa_2$
,,	L	$\gamma_2\lambda_2$
IgM	K	$(\mu_2\kappa_2)_6$
,,	L	$(\mu_2\lambda_2)_6$
IgA	K	$(\alpha_2\kappa_2)_n$ (n = 1, 2, or ?3)
,,	L	$(\alpha_2\lambda_2)_n$
IgD	K	$\delta_2\kappa_2$
,,	L	$\delta_2\lambda_2$

Since rheumatoid factor reacts only with appropriately treated immunoglobulins (human or animal) of the IgG class, it may now be inferred that part, at least, of the (heavy) γ-chain is involved, since the (light) κ and λ chains of the IgG are also present in immunoglobulins of other classes. Before describing tests to confirm this inference, it is interesting to consider the effects of proteolytic digestion of immunoglobulins.

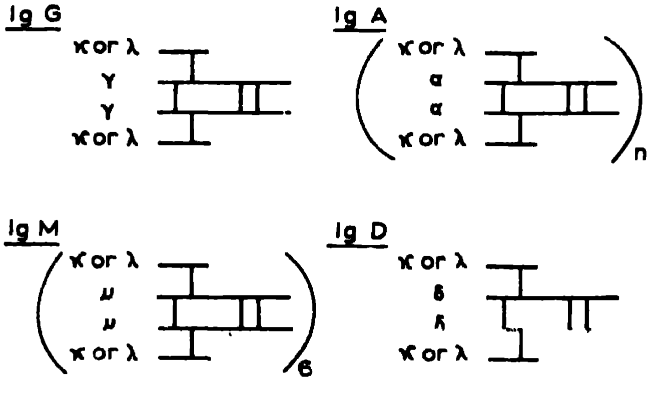


FIG. 2. Structure of the four immunoglobulins. In this diagram, the designations are those now adopted for the 'heavy' and 'light' polypeptide chains. IgM and IgA are regarded as polymers, but otherwise conform to the general immunoglobulin structure.

Proteolysis was, of course, the first method used for controlled degradation of immunoglobulins, and the more readily available IgG of animal origin was tested initially. Human immunoglobulins are, however, of special interest, even though there are certain unresolved difficulties with these preparations. Proteolysis has been achieved with various enzymes, particularly with papain or with pepsin; again, the particular technical differences and difficulties of the various procedures need not obstruct our broad understanding of the present position. The principal action of proteolytic enzymes is to break down that part of the 'heavy' polypeptide chain which is not attached to the light chains (see Fig. 1). Thus a fragment is detached from human IgG which can be distinguished by its 'fast' mobility on electrophoresis from the 'slow' remaining part or parts of the immunoglobulin. This 'fast' moiety obtained by papain digestion is to be known as the Fc-fragment. This fragment from rabbit IgG can be *c*rystallized, and this fragment also contains most of the immunoglobulin *c*arbohydrate (either of these properties serving as useful mnemonics for the designation, F*c*). Naturally, this fragment derived from the heavy chain varies in its nature according to the immunoglobulin class from which it is made; the new notation allows for this distinction as Fcγ, Fcα, etc. The slow fragment, when intact, contains the whole of two light (κ- or λ-) chains and part of two heavy chains; in this form it is to be known as $F(ab)_2$. With papain, however, there is usually some further cleavage of this slow-moving material into two more or less equal-sized parts, known as Fab-fragments. The '-ab' in this designation serves to represent 'antibody', for it is this part of the immunoglobulin molecule (and this part alone) which retains this special characteristic. As Fab, it is usually a univalent, non-precipitating antibody, but specific precipitation with antigen may, under appropriate circumstances, be obtained with $F(ab)_2$. This cleavage into fragments is illustrated in Fig. 3. Further treatment of digests containing Fab fragments with reducing agents liberates light (κ- or λ-) chains, and in addition the residual part of the heavy chain, which is to be known as Fd.

Whereas light and heavy chains (e.g. κ- and λ-chains) can be obtained in consistent and reproducible forms, the products of proteolysis are less homogeneous. This is particularly the case with Fab-fragments, some of which in any preparation are apparently rather more highly charged than the corresponding fast Fc-fragment, others being of lower charge. This inconstancy is of some practical consequence, since while at present it is reasonably easy to obtain Fab-fragments free from Fc (e.g. by preparative electrophoresis), Fc-fragments free from Fab cannot usually be obtained. In theory, of course, a specific antiserum absorbed with κ- and λ-chains might be used to remove Fc-fragments, but this has not yet been achieved.

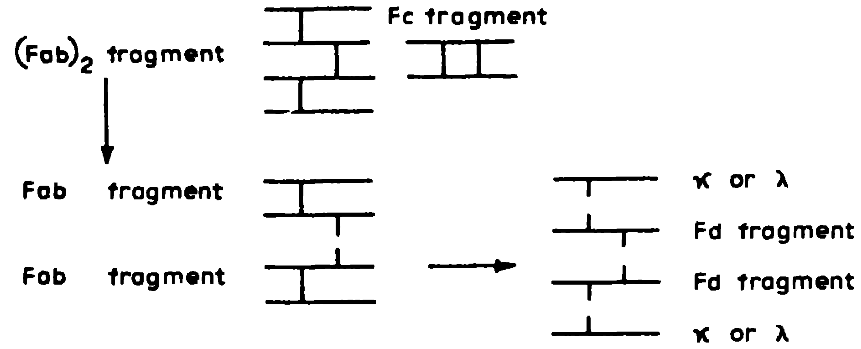


FIG. 3. This diagram represents the results of proteolytic digestion and of subsequent reduction. It also shows the designation of each fragment.

It has been stated above that rheumatoid factor reacts with the heavy (i.e. γ-) chains of IgG. From recent studies of the digestion products of IgG (Henney and Stanworth, 1964), there seems to be little doubt that the fast (Fc) fragments contain the reactive sites. These authors cautiously point out that their experiments, involving quantitative estimations of precipitable nitrogen from various mixtures, were all preceded by heat aggregation of partially separated fragments; in theory, therefore, inadequate aggregation of the slow (Fab) fragments might have been the underlying reason for the comparatively small amounts of precipitate harvested from admixtures of rheumatoid factor. Nevertheless, the amount of precipitate given by the Fc-fragments was such that, within the limits of experimental error, the whole amount of IgG-rheumatoid factor precipitate could be accounted for on the basis of reactivity being confined to the Fc moiety of the IgG.

Human Allotypes

At this point, the relevance of the discoveries of the Gm grouping system of human immunoglobulin allotypes must be pointed out, though the current situation is succinctly set out in Glynn's review (1963). This system was discovered when individual human antisera of incomplete Rhesus type were adsorbed on to appropriate human red cells, these coated cells being substituted for the coated sheep cells as used in the Waaler-Rose test. With this individual Rhesus system, there were wide variations between different individuals' specimens of rheumatoid serum in their ability to cause agglutination. Furthermore, the result with those systems using rheumatoid sera which did cause satisfactory agglutination could be inhibited by some specimens of normal IgG, but not by others. On this basis a series of Gm factors have been defined, depending on specific antigens located on the Fc-fragment of human IgG. This work gives more evidence that the IgG site which combines with rheumatoid factor is similarly confined mainly, if not entirely, on the Fc fragment. This site, it is thought, must be exposed by some mild form of denaturation for IgG samples to be fully reactive; the precise nature of this mild degradation which can be brought about *in vitro* is being determined by ultracentrifuge and optical rotatory dispersion studies in parallel with the biological tests (Henney and Stanworth, 1965). Much remains to be done before this 'antigen exposure' process is fully understood, but enough information is now available to encourage renewed speculation about the kind of process involved in the body of the patient with rheumatoid arthritis. Firstly, however, a few facts should be given about the nature of rheumatoid factor itself.

Nature of Rheumatoid Factor

The evidence for regarding rheumatoid factor as IgM is part physico-chemical, part immunological. Rheumatoid factor activity can be concentrated in a variety of ways (cf. James, Felix-Davies and Stanworth, 1961). In the preparative ultracentrifuge, separation of the 19S class removes most of the activity; this class of high molecular weight proteins is mainly composed of IgM and α_2-macroglobulin (high molecular lipoproteins may be separated in other ways without loss of activity). Similarly, on Sephadex G.200 (Killander, 1964) the rheumatoid factor titre per mg of protein is greatest in the leading fractions, again mostly IgM and α_2-macroglobulin if lipoprotein has been removed. Precipitation of euglobulin by addition of 14 parts of water to 1 part of serum brings down rheumatoid factor and IgM in the precipitate. Such processes can be used to concentrate rheumatoid factor and give increasing potencies per mg protein, though with some techniques, especially those using charged cellulose columns and particularly when the IgM has been largely separated from other proteins, yields tend to fall, indicating destruction of biological activity. Analytical ultracentrifuge tests on whole serum specimens with potent rheumatoid activity may show some elevation of 19S class proteins, but in view of the varying amounts of α_2-macroglobulin in this fraction and of the very small absolute quantities of rheumatoid factor involved even in very high titre sera, this finding is in no way diagnostic. Additional peaks of higher S-constant (e.g. 22-S) are sometimes seen, and these probably depend on combinations of rheumatoid factor IgM with IgG, but this phenomenon, according to Svartz (1960), may well be due to some *in vitro* denaturation and should not necessarily be regarded as of biological significance.

The ability of IgG-coated erythrocytes in the Waaler-Rose reaction to adsorb rheumatoid factor has been inferred from the agglutination phenomenon, in which adjacent erythrocytes are presumably bridged by multivalent IgM. Some indications of the amounts attached to the cell surfaces have been given by experiments (Greenbury, 1960*b*) using sera radioactively labelled *in vitro*. But proof of the nature of the protein involved has been obtained by Rowe (1962) who used dilutions of rheumatoid factor just insufficient of themselves to cause full agglutination. Agglutination could then be effected by admixture of rabbit antibodies directed specifically against certain immunoglobulin fragments; using modern terminology for these, the effect was observed with sera containing 'anti-Fab', and with 'anti-μ-chain', but not with sera containing only 'anti-γ-chain'. These tests seem to show that the only immunoglobulin adsorbed by the sensitized sheep cells was IgM, and come close to proving that rheumatoid factor is indeed of this nature. The immunoglobulin, IgM, is known to carry antibody activity in a number of familiar systems. It is, of course, associated with the naturally occurring isoagglutinins involved in ABO grouping, this kind of activity being apparently stimulated in persons of the appropriate hereditary make-up by foreign antigens (probably of bacterial origin) which resemble the blood group-specific polysaccharide haptens. IgM is concerned with complement-fixing activity involved in the Wassermann reaction for syphilis, and, in most instances, with somatic (anti-O) typhoid antibodies. In a number of these instances, the IgM antibody titre persists for lengthy periods. On the other hand, various lines of investigation (cf. Turner and Rowe, 1964) suggest that IgM antibody activity is an early phase of the immune response to some antigens of bacterial origin (e.g. anti-H typhoid), and is normally succeeded by antibody formation in the form of IgG immunoglobulins. It is too soon as yet to define the underlying basis of the biological significance of these differences. Certainly, one cannot conclude at present that persistence of IgM antibody activity is a feature peculiar to rheumatoid factor or that it necessarily implies continued release of antigen. Since rheumatoid factor is a modified immunoglobulin, it may well be an antibody. If so, exactly what is it directed against, and what is the stimulus for its production? If it is to be regarded as an 'autoantibody', a strict definition would require that immunoglobulin production was being directed against a normal body constituent. Yet, except for the 22S conjugate detectable in a minority of rheumatoid sera (and regarded by some as an *in vitro* artefact), rheumatoid factor reactions do not seem to fall into this category. Gm-grouping studies by Fudenberg and Kunkel (1961) if confirmed, indicate that the antibody activity of the individual rheumatoid serum is never in fact directed against the Gm group demonstrable in that individual's IgG; thus, the Gm-type rheumatoid antibodies are iso-antibodies—that is they are directed against groupings present in other individuals of the same species who differ in allotype.

This line of argument would put rheumatoid factor into the iso-antibody category if it were all of the Gm-type. In fact, though the Gm-type antibodies are principally in the IgM class, they do not appear to constitute a large proportion of the total rheumatoid factor in most patients' sera. Repeated absorption with a Gm cell system, for example, does not appreciably reduce the titre of the Waaler-Rose reaction. It may be concluded that Gm-type antibodies are one, usually minor, component of a range of 'rheumatoid factors' which vary in a single specimen of serum in their particular specificities.

The main Waaler-Rose type rheumatoid factor is strictly speaking neither an 'auto-antibody' nor an 'iso-antibody'. Instead, it appears to be an antibody directed against various groupings on the Fc-moiety of any human IgG (present also in varying degree in the IgG of various mammalian species), which, as already mentioned, can be disclosed by minor degrees of 'unfolding' of the immunoglobulin molecule. At present, it seems most reasonable to conclude that the original antigen exciting this antibody formation is found in those places where plasma cells abound in the rheumatoid patient, e.g. in the joints. Whether the IgG antigen is in the altered 'unfolded' form as the result of faulty production by other plasma cells poisoned or damaged in some way, or whether the IgG, normal at the time of its release from its parent cells, is subsequently distorted remains to be determined. In the latter case, the distortion might be effected by free enzymes in the inflamed area or by some process such as an antigen-antibody combination, e.g. with some self-multiplying infective agent. Posed in this form the chief problem would seem to be why some form of anti-IgG antibody is not, apparently, formed in all ordinary persons whenever their own induced IgG antibodies combine within the body with some foreign organism or other protein. Though still speculative in relation to human disease, it is now known that in animals IgG antibody which has reacted with antigen, can, on re-injection, give rise to 'anti-antibody' which is truly specific.

SIGNIFICANCE OF RHEUMATOID FACTOR TESTS

From these leads, which suggest many kinds of further investigation on the fluids and tissues of rheumatoid patients, it may be worthwhile to return to the test itself and consider its significance and possible improvement. The presence of rheumatoid factor in strong titre, of course, reinforces a diagnosis of rheumatoid arthritis, and the consistent absence of the factor in such disease groups as ankylosing spondylitis, psoriatic arthritis, etc. certainly seems to point to differences in pathogenesis. The existence of a proportion of apparently typical rheumatoid patients with negative rheumatoid factor tests presents a continuing puzzle which merits further investigation.

Sufficient authorities are of the opinion that extremely high titres of rheumatoid factor are of serious prognostic significance in relation to the more striking complications of rheumatoid disease. Nor do titres fall with most of the treatments regarded as being beneficial for alleviation of the disease; unquestionably, short-term changes in activity of the disease process are to be studied not in terms of rheumatoid factor activity, but with the erythrocyte sedimentation rate, C-reactive protein tests—now available in quantitative form (Crockson, 1963)—or of other 'acute-phase protein' changes.

As currently performed, the Waaler-Rose reaction is not truly quantitative as between one laboratory and another; indeed, probably in many laboratories, the absolute titre is not reliable from year to year. Such facts vitiate many of the current views on clinico-pathological correlations. It is therefore good to know that the Medical Research Council has started a search for a suitable standard rheumatoid factor preparation which can be issued for reference purposes. The studies of this group have already shown (Bangham, personal communication) that in relation to a standard, different laboratories testing exchanged specimens give accurately reproducible relative findings.

Though the Waaler-Rose reaction can undoubtedly be refined in this way, it seems likely that it is due for replacement within a few years at least by some test based on a reproducible human pooled IgG derivative. In view of the difficulties of controlled aggregation of the γ-globulin fragments to produce antigens and of the uncertainties of agglutination and even of precipitation end-points in the test itself, it may be predicted that an inhibition form of test will eventually be used. Suitable inhibitors might consist of stabilized γ-chains derived by reduction of IgG and freed from κ- and λ-chains, or perhaps of standard Fcγ fragments. Theoretically, also, the titres found might be less subject to deviation by inhibitory substances, naturally present in many rheumatoid sera, if the IgM were at least partially purified, as in many current research programmes. Meanwhile, there is still much to be learned from the study of carefully conducted Waaler-Rose tests with the inclusion of the soon-to-be-available standard.

ACKNOWLEDGEMENTS

My colleagues in the Department of Experimental Pathology have generously given me much help in the preparation of this account. Their researches have been supported by the Nuffield Foundation, the Wellcome Trust, the Sumner Fund, the Rockefeller Foundation, the Medical Research Council and the Arthritis and Rheumatism Council.

REFERENCES

BANGHAM, D. R. Personal communication.

BULL. WLD. HLTH. ORG. (1964) **30**, 447.

CROCKSON, R. A. (1963) *J. clin. Path.*, **16**, 287.

EDELMAN, G. M., KUNKEL, H. G. and FRANKLIN, E. C. (1958) *J. exp. Med.*, **108**, 105.

FUDENBERG, H. H. and KUNKEL, H. G. (1961) *J. exp. Med.*, **114**, 257.

GLYNN, L. E. (1963) 'Rheumatoid Arthritis'. Ch. 25 in *Clinical Aspects of Immunology*, ed. Gell, P. G. H. and Coombs, R. R. A. Blackwell, Oxford.

GREENBURY, C. L. (1960*a*) *J. clin. Path.*, **13**, 325.

GREENBURY, C. L. (1960*b*) Empire Rheumatism Council Symposium 'Immunology in the Rheumatic Diseases', p. 33.

HENNEY, C. S. and STANWORTH, D. R. (1964) *Nature* (*Lond.*), **201**, 511.

HENNEY, C. S. and STANWORTH, D. R. (1965) *Immunology*, **9**, 139.

JAMES, K., FELIX-DAVIES, D. and STANWORTH, D. R. (1961) *Ann. rheum. Dis.*, **20**, 369.

JAMES, K., STANWORTH, D. R. and HENNEY, C. S. (1964) *Nature* (*Lond.*), **202**, 563.

KILLANDER, J. (1964) *Acta path. microbiol. scand.*, **61**, 127.

PORTER, R. R. (1959) *Biochem. J.*, **73**, 119.

ROSE, H. M., RAGAN, C., PEARCE, E. and LIPMAN, M. O. (1948) *Proc. Soc. exp. Biol.* (N.Y.), **68**, 1.

ROWE, D. S. (1962) *Immunology*, **5**, 549.

ROWE, D. S. (1964) 'Rheumatoid Factors'. Ch. 22 in *Recent Advances in Clinical Pathology*, ed. Dyke, S. C. Churchill, London.

ROWE, D. S. and FAHEY, J. L. (1965) *J. exp. Med.*, **121**, 185.

SQUIRE, J. R. (1964) *Proc. roy. Soc. Med.*, **57**, 749.

SVARTZ, N. (1960) *Acta med. scand.*, **168**, 285.

TURNER, M. W. and ROWE, D. S. (1964) *Immunology*, **7**, 639.

WAALER, E. (1940) *Acta path. microbiol. scand.*, **17**, 172.

IX

Histochemical Techniques in the Study of the Cell Biology of Dental Tissues

A. R. TEN CATE

HISTOCHEMISTRY, that is the study of the chemistry of cells and tissues as opposed to their morphology, is a widely used technique for obtaining information about the activities of cells and tissues. The development and function of the dental tissue involve a spectrum of cell activity ranging from fibrillogenesis to the production of the hardest mineralized tissue in the body and, as such, should be ideal for study with histochemical methods. Unfortunately, theory and practice do not always coincide. One of the first principles of histochemistry is to avoid any change in the chemical architecture of the cell or tissue under investigation during the preparation of sections for microscopical examination. Demineralization in acid, a normal procedure in the preparation of sections of dental tissue, is hardly compatible with this histochemical principle and therefore alternative methods of section preparation must be used.

THE PREPARATION OF SECTIONS OF DENTAL TISSUE SUITABLE FOR HISTOCHEMICAL INVESTIGATION

Before discussing the various ways of preparing dental tissue for histochemical investigation mention must be made of a recent technical advance which has added considerable impetus to the practice of histochemistry; the development of the cryostat. This is simply a refrigerated microtome operating at a temperature of $-15°$ C. to $-20°$ C. and it enables the cutting of thin, less than 5 μ if desired, frozen serial sections. As a general rule such sections are preferred in any histochemical investigation as their chemistry approximates most closely to that existing *in vivo*.

(1) *Sectioning of dental tissue without demineralization*

If developing teeth are used, it is possible to cut sections without having first to decalcify the hard tissue. Plate V, Fig. 1 is of a human tooth germ which has been sectioned in this way and in which a considerable amount of hard tissue is present.[1] Several cellular activities can be studied in such sections; fibrillogenesis in the pulp, mineralization in both the dentine and enamel and also the functioning odontoblast and ameloblast.

This approach can be extended to the use of teeth in which development is almost, but not quite, complete. Such teeth have erupted into the mouth and are functional, but their roots are not yet fully formed. Transverse sections cut on a cryostat through the developing root region retain a surprising amount of hard tissue in good apposition with the soft tissue (Plate V, Fig. 2) and using such sections, it is possible to study histochemically the organizing root sheath epithelium, dentinogenesis and cementogenesis.

Less success has been achieved with attempts to section mature teeth without prior demineralization. Most dental laboratories have a 'tooth slicing' machine somewhere on the premises, often self-designed and self-made. Modifications of such machines, usually involving the incorporation of elaborate cooling devices, have been used from time to time to prepare sections of mature teeth suitable for histochemical study. The resultant sections are thick and the soft tissue is usually destroyed; even so Symons (1962), using this approach, was able to demonstrate alkaline phosphatase activity in the odontoblast process.

(2) *Sectioning of dental tissue with demineralization*

Some fifteen years ago several attempts were made to demineralize dental tissue and at the same time preserve its chemical architecture. A limited success was gained by using buffered mineral acids to retain activity of the relatively stable enzyme alkaline phosphatase during demineralization. A great deal of enzyme activity however is destroyed in the acid and, as the intracellular localization of enzyme activity was poor, this approach was soon abandoned. In 1962 Balogh Jr. reported good preservation of oxidative enzyme activity in partially demineralized dental tissue. An organic chelating agent, ethylenediaminetetraacetic acid (EDTA) was used for a short period to demineralize partially the tissue which was then frozen and sectioned in a cryostat. As the oxidative enzymes are confined to the mitochondria, which are stable in EDTA, the problem of diffusion of enzyme during the period of demineralization did not arise.

---

[1] The plates referred to in this paper will be found between pages 154–5.

(3) *Bulk staining*

Provided that the final reaction product of a histochemical method is insoluble in acid, the method can be applied to small pieces of dental tissue, such as for example small fragments of dentine with attached pulpal tissue, and, after the reaction has been completed, the specimen can be decalcified in acid and sectioned. In this way it is possible to demonstrate hydrolytic enzyme activity in mature odontoblasts using the azo dye techniques as these dyes are insoluble in acid (Osborne, 1965).

(4) *Isolation of the soft dental tissues*

So far as the study of cell activity is concerned, it is no great handicap to investigate the soft dental tissues divorced from their *in-vivo* position. Once isolated from the hard tissue elements, unfixed frozen sections of the soft tissue can be readily prepared for study with histochemical methods. Most, if not all of the dental soft tissues can be obtained for histochemical investigation if this approach is used.

By using one or several of the above methods most of the dental tissues can be examined histochemically. A review of all the histochemical methods which are available at the present time would be a fruitless exercise as these have been well documented in Pearse's *Histochemistry* (1960). However, it is rewarding to trace the evolution of methods to demonstrate one particular cell constituent, the enzyme acid phosphatase, as many of the principles and pitfalls of histochemistry are illustrated.

METHODS FOR THE DEMONSTRATION OF ACID PHOSPHATASE ACTIVITY

Gomori (1941) first devised a technique for the demonstration of acid phosphatase activity which depends upon the hydrolysis of phosphate ions from an organic ester by the enzyme. The released phosphate ions are permitted to combine with lead to form a precipitate of lead phosphate at the site of enzyme activity. This deposited lead phosphate is, in turn, converted to a visible precipitate of lead sulphide by treatment with ammonium sulphide. At that time a fundamental principle was that fixation was essential for the accurate intracellular demonstration of enzyme activity. Adherence to this caused some degree of schizophrenia as a further histochemical principle needed to be observed, the preservation of the maximum amount of enzyme activity.

Chilled acetone was selected as the fixative of choice but unfortunately, this largely inactivated an extremely labile enzyme. This meant that to produce a positive result long incubation times in the substrate solution were necessary, greatly enhancing the possibility of enzyme diffusion within the cell. Non-specific deposition of lead salts tended to occur and there was also the possibility of diffusion of the precipitated lead phosphate to consider. Because of these hazards an alternative technique was sought for and found. This technique depended upon the principle that azo dyes form a coloured reaction product when combined with a naphthol group. If a substrate, which on hydrolysis reveals a free naphthol group, is exposed to the enzyme the naphthol group combines instantly with an azo dye in the substrate solution to produce a coloured reaction product at the site of enzyme activity. This simultaneous coupling method as it is called, eliminates the possibility of diffusion of the reaction products which occur with the Gomori method. As diffusion was restricted it was thought that, after acetone fixation, accurate intracellular localization of enzyme activity was being obtained. The resulting picture was of dye distributed throughout the cytoplasm of the cell. The significant step in the understanding of acid phosphatase localization within cells came in 1954 when it was demonstrated biochemically by de Duve that acid phosphatase activity, and activity of other hydrolytic enzymes, was confined to small cytoplasmic particles with a limiting phospholipid membrane termed lysosomes. It became obvious that fixation in acetone to preserve accurate intracellular localization was, in fact, exerting the opposite effect. The acetone caused disruption of the retaining envelope and permitted the contained enzymes to diffuse into the cytoplasm of the cell where they were demonstrated histochemically. The wheel has now turned the full circle and, for the demonstration of acid phosphatase activity within cells, the Gomori lead nitrate technique, with certain modifications, is used once more. To preserve the integrity of the lysosomes the solutions used in the Gomori technique are made isotonic and applied to unfixed frozen sections or to sections fixed in formol calcium. The resulting picture is of particulate deposits of lead sulphide indicating the lysosome location of enzyme activity instead of a diffuse cytoplasmic deposition (Plate VI, Fig. 3).

It is evident, from this brief account of the evolution of methods to demonstrate acid phosphatase activity, that there are still many pitfalls in the practice of histochemistry. Some of these can be avoided by strict adherence to principles; others may not be apparent at this time and may be revealed when further information is available.

To illustrate the information gained from the use of histochemistry in the study of the function of the dental tissues three particular problems will be discussed. These are first the function of the dental epithelium during tooth eruption and its role in the establishment of the epithelial attachment; second, the function of the epithelial debris of Malassez in the periodontal ligament and third, the problem of dentine sensitivity.

INVESTIGATION OF THE DENTAL EPITHELIUM DURING TOOTH ERUPTION AND ESTABLISHMENT OF THE EPITHELIAL ATTACHMENT

After the completion of amelogenesis it has been suggested that the dental epithelium degenerates (Diamond and Weinmann, 1940). On the other hand it has also been suggested that it protects the completed enamel surface from resorption by the surrounding follicular tissue; causes dissolution of this connective tissue during eruption of the tooth: provides, with oral epithelium, an epithelium lined pathway for the tooth to erupt through; and finally assists, together with oral epithelium, in the histogenesis of the epithelium of attachment (McHugh, 1961) Histochemical investigation (Ten Cate, 1963) has shown that both views are in part correct. A method was required for demonstrating the functional activity of the dental epithelium Fortunately, at the time this problem was being considered, the concept of the lysosome was introduced and it was demonstrated that consistent and reproducible changes occurred in the lysosome content of the cell as it degenerates. It has been demonstrated that this change occurs in 'organized' cell death which takes place during development, for example in the degenerating tadpole tail (Novikoff, 1961) and in the degenerating Mullerian duct (Brachet, Decrolly-Briers and Hoyez, 1958). Therefore, if lyosomes exist in dental epithelium, these can be used as natural markers to indicate the viability of the epithelium. Lysosomes were demonstrated in the dental epithelium and their configuration indicated that over the occlusal surface of the unerupted tooth the ameloblast layer of the reduced dental epithelium was degenerating whereas the remaining epithelial cells of the dental epithelium retained their viability. On the other hand, on the evidence of their lysosome pattern, over the non-occlusal surface of the enamel all components of the dental epithelium were found to be degenerating. Furthermore it was also possible to demonstrate basal oral epithelium cells on the connective tissue surface of the epithelium in this location.

Oral epithelium cells could be distinguished from dental epithelium by the fact that alkaline phosphatase activity could be demonstrated in the latter but not in the former. These findings indicate a differing functional role for the dental epithelium depending upon its location. Over the occlusal surface of the erupting tooth the connective tissue intervening between the tooth and the oral mucosa has to be destroyed and, although it is not absolutely certain that this destruction is brought about by the dental epithelium, it seems highly likely. In any case the dental epithelium in this location needs to retain its viability so that it can fuse with the oral epithelium and form an epithelial lined pathway through which the tooth erupts. On the other hand the dental epithelium covering the non-occlusal surface of the crown has no need to exert a desmolytic effect on the follicular connective tissue. In this region a strong attachment between the tooth and surrounding gum is required. From the histochemical evidence the sequence of events would seem to be that the dental epithelium degenerates and, because of this, there is a migration of viable basal oral epithelium cells over the degenerating dental epithelium which ultimately establish the epithelium of the attachment.

Investigation of the epithelial debris of Malassez

After eruption of the tooth, remnants of dental epithelium persist in the periodontal ligament as the epithelial debris of Malassez. In this location these epithelial rests persist for seventy years or more and are apparently functionless unless infection occurs, when the epithelium proliferates to form the lining of dental cysts. To determine if these epithelial cell rests had any function or, if not, why this epithelium persisted for such long periods of time, their oxidative enzyme content was investigated.

One of the first findings showed that the epithelial cell rests contained glycogen and it seemed that an analysis of the glycolytic pathways might provide information on the metabolic activity of the cell rests and give a clue to their function. The demonstration of oxidative enzymes in tissue sections depends upon the fact that colourless tetrazolium salts when reduced form a coloured formazan salt. The tetrazolium salt is used to accept the electrons from the oxidized substrate which, if the tetrazolium salt was not present, would pass on by a series of successive electron carriers to water. In this way, depending upon the substrate available for oxidation, a whole series of oxidative enzymes can be demonstrated. In the epithelial cell rests succinic dehydrogenase activity, lactic dehydrogenase activity, glucose-6-phosphate dehydrogenase activity and the activities of reduced triphosphopyridinenucleotide and reduced diphosphopyridinenucleotide were demonstrated (Plate VI, Fig. 4), the latter two being indicative of co-enzyme I and co-enzyme II activity. The significance of these findings has been discussed elsewhere (Ten Cate, 1965) but in summary, the presence of glycogen in the epithelial cell rests together with a functioning pentose shunt and the presence of an anaerobic glycolytic pathway is indicative of a metabolism requiring little energy and suggests that the epithelium has no functional role.

Investigation of dentine sensitivity

In 1959 Avery and Rapp claimed to have demonstrated the activity of acetylcholinesterase in the nerves of the pulp; in a localized band of pulp underlying the pre-dentine of the crown of the tooth; and throughout the odontoblast and its process. They concluded that pain conduction in dentine was transmitted via the odontoblast process to the odontoblast cell body where synapses existed between the free nerve endings of the pulp and the odontoblast thereby completing a neuro-anatomical pathway. The technique they used was that of Koelle and Friedenwald (1949) which depends upon the hydrolysis of acetylcholine by acetylcholinesterase in the presence of copper. A precipitate of copper thiocholine is formed at the site of enzyme activity which is converted to a visible product, copper sulphide, by washing with ammonium sulphide. Many criticisms have been levelled at this technique of which the most severe is that the conversion of copper thiocholine involves its dissolution and replacement by an amorphous deposit of copper sulphide. In other words there is not a direct replacement *in situ* and the copper sulphide is not topically related to the original crystals of copper thiocholine. Localization cannot therefore be regarded as cytologically accurate by this method. Avery and Rapp (1959) were aware of these criticisms and tried to account for them but they were at a further disadvantage in that they had to use thick sections and therefore long incubation periods with the attendant possibility of diffusion. In 1964 Karnovsky and Roots devised a 'direct' method for the demonstration of acetylcholinesterase which eliminated the substitution step. This depends upon the fact that liberated thiocholine reduces ferricyanide to ferrocyanide which in the presence of copper forms an insoluble, finely granular precipitate of copper ferrocyanide at the sites of enzyme activity. It seemed worthwhile to repeat the investigation of Avery and Rapp using this more precise technique and thinner sections. Initially transverse, unfixed frozen sections through the developing root region were used. In such sections, after three hours' incubation in the substrate medium acetylcholinesterase activity was precisely located to nerve bundles entering the apical foramen and also to non-myelinated nerve fibres (Plate VII, Fig. 5). No enzyme activity could be demonstrated either in the odontoblasts or in the zone of pulp tissue immediately below the odontoblasts. It was of interest that periodontal ligament tissue on the same section showed a great deal of acetylcholinesterase activity. To counter the possibility that the incubation period was too short, and therefore did not allow demonstration of minimal amounts of acetylcholinesterase activity, sections were also incubated for 24 hours. The results were the same as after a three-hour incubation period except that the dye was deposited in greater amounts in the reactive sites. These sections were taken from the developing root region and there is evidence that the dentine here is not innervated (Fearnhead, 1959); possibly because the neuro-anatomical pathway, if it exists, may not yet have been established. It was therefore necessary to examine coronal dentine for evidence of acetylcholinesterase activity. This posed the difficult technical problem of how to obtain unfixed frozen sections of coronal odontoblasts and coronal dentine.

Several approaches were tried. First, extracted teeth were split open and their pulps removed, frozen and sectioned. No odontoblasts were present in such sections: however, nerve trunks ascending in the pulp and occasional fine non-myelinated nerve fibres showed acetylcholinesterase activity (Plate VII, Fig. 6). At the periphery of the section, which most likely consisted of the pulp immediately below the odontoblasts, there was no reaction. As there were no odontoblasts in these sections the next step was to scrape the pulpal wall of the dentine after removal of the pulp and to prepare a smear of the scraping. On the same slide a teased pulpal nerve bundle was placed to act as a control. In such a smear isolated small fragments of dentine were identified but it was not possible to identify with absolute certainty the cellular elements as odontoblasts. No staining reaction was produced in such smears except in the teased nerve

PLATE V

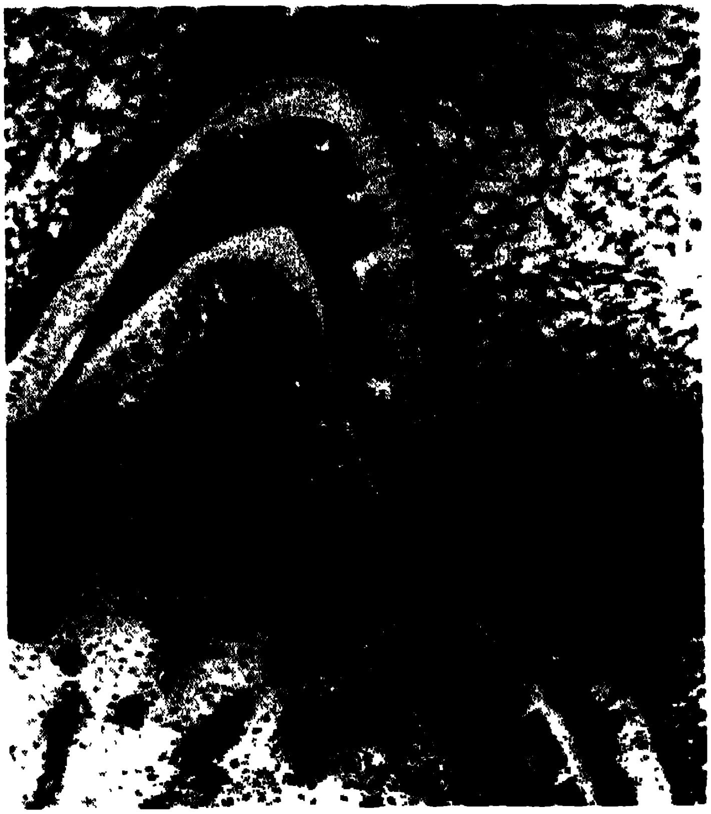


FIG. 1. Photomicrograph of a human developing tooth germ sectioned without demineralisation.

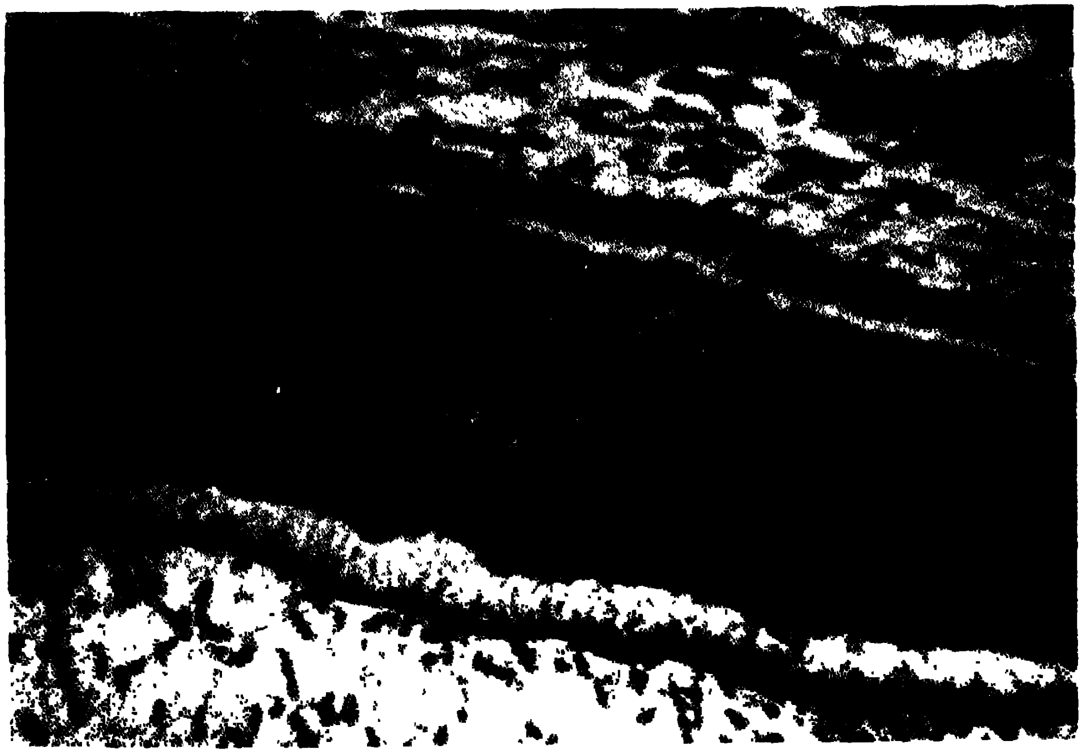


FIG. 2. Photomicrograph of a frozen transverse section through the forming root region of a tooth illustrating undemineralised dentine (*a*) and cementum (*b*) in section.

PLATE VI

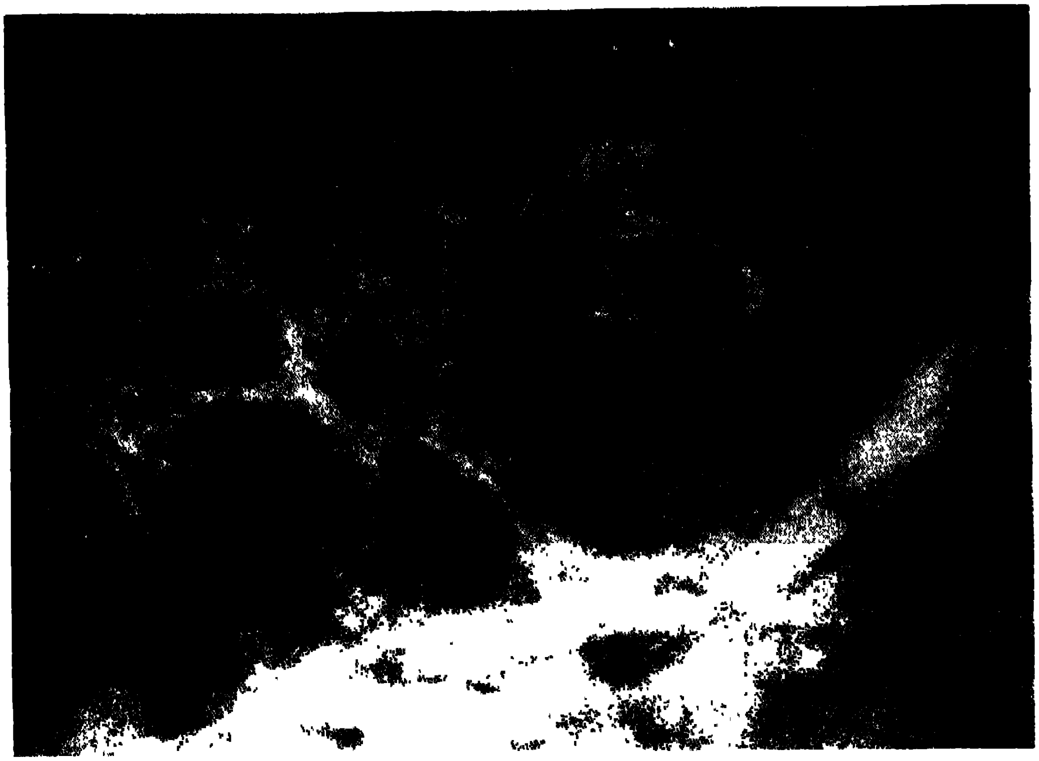


FIG. 3. Frozen section of dental epithelium illustrating the particulate localization of acid phosphatase activity after application of the Gomori lead nitrate technique.

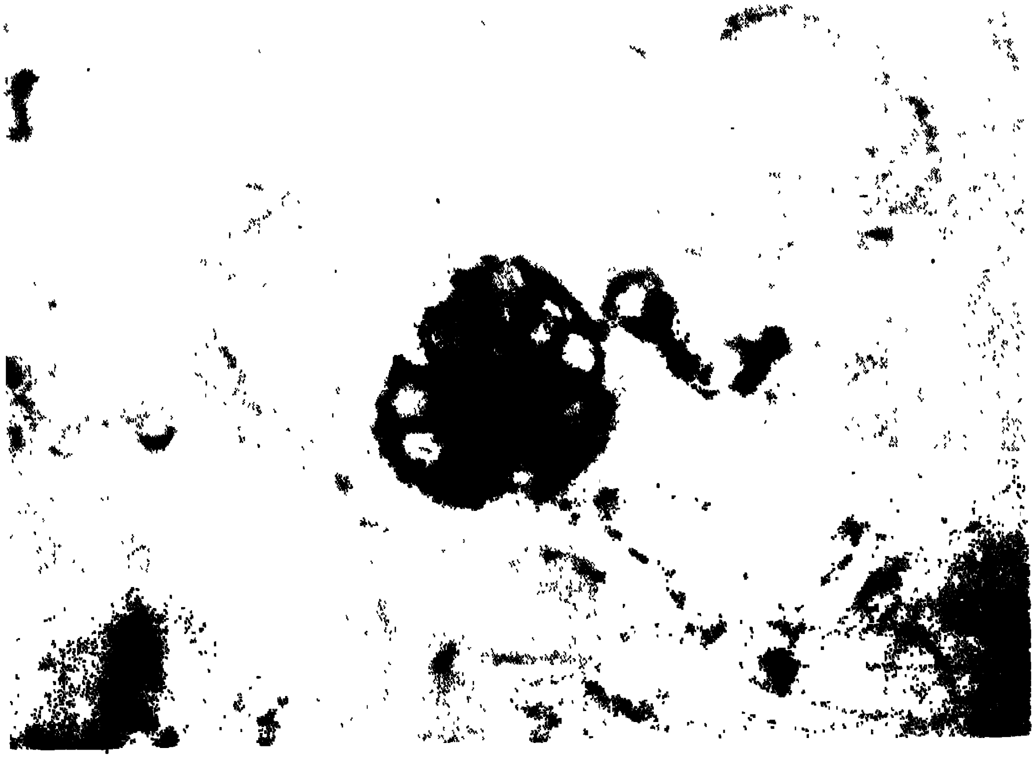


FIG. 4. Diphosphopyridine nucleotide diaphorase activity in an epithelial cell rest. Enzyme activity indicated by the particulate deposits of formazan.

PLATE VII

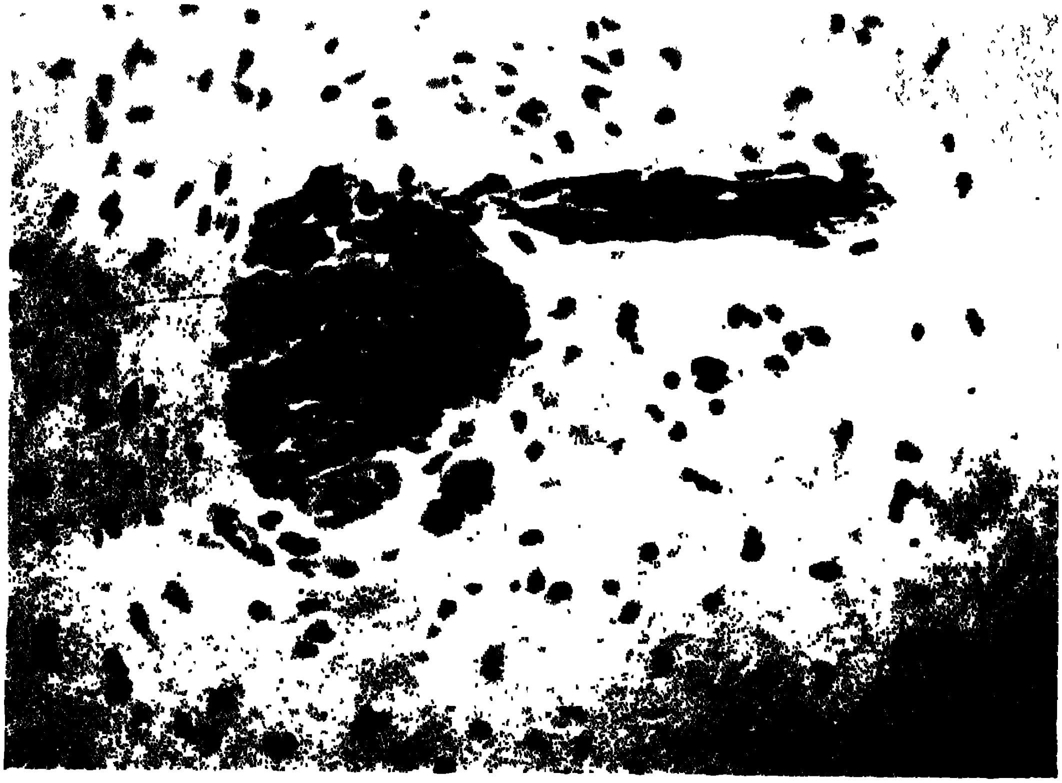


FIG. 5. Transverse section through the pulp in the root region illustrating the localization of acetylcholinesterase activity in a nerve bundle.

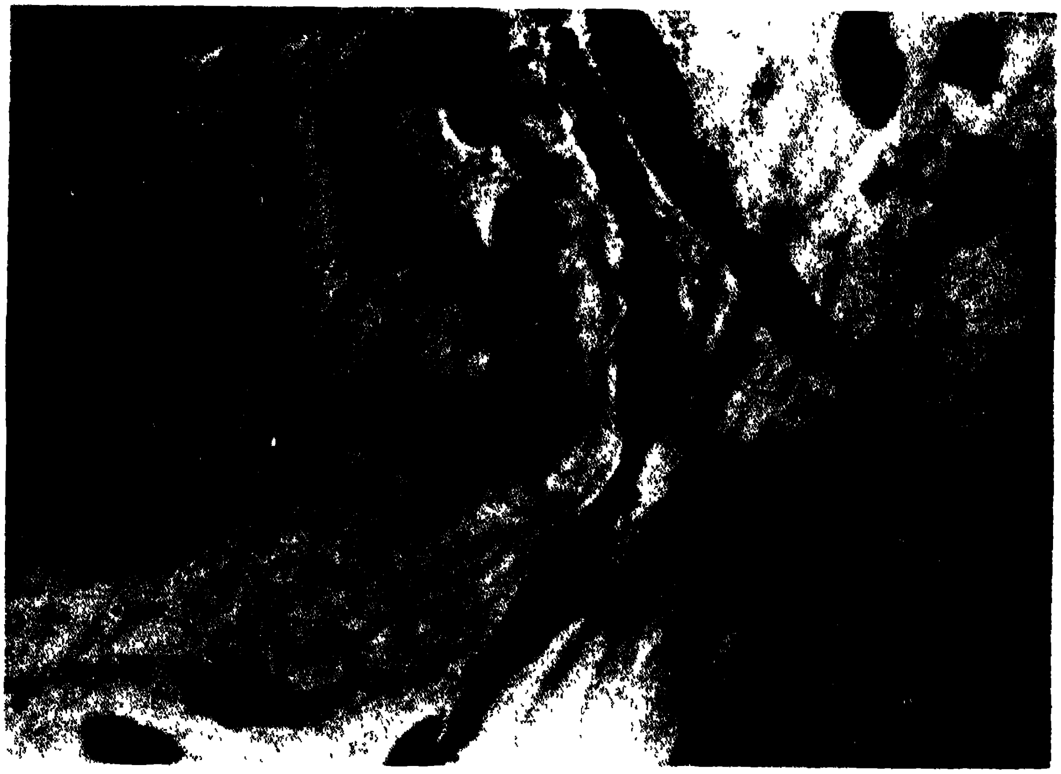


FIG. 6. Acetylcholinesterase activity in a fine nerve filament in the pulp.

PLATE VIII

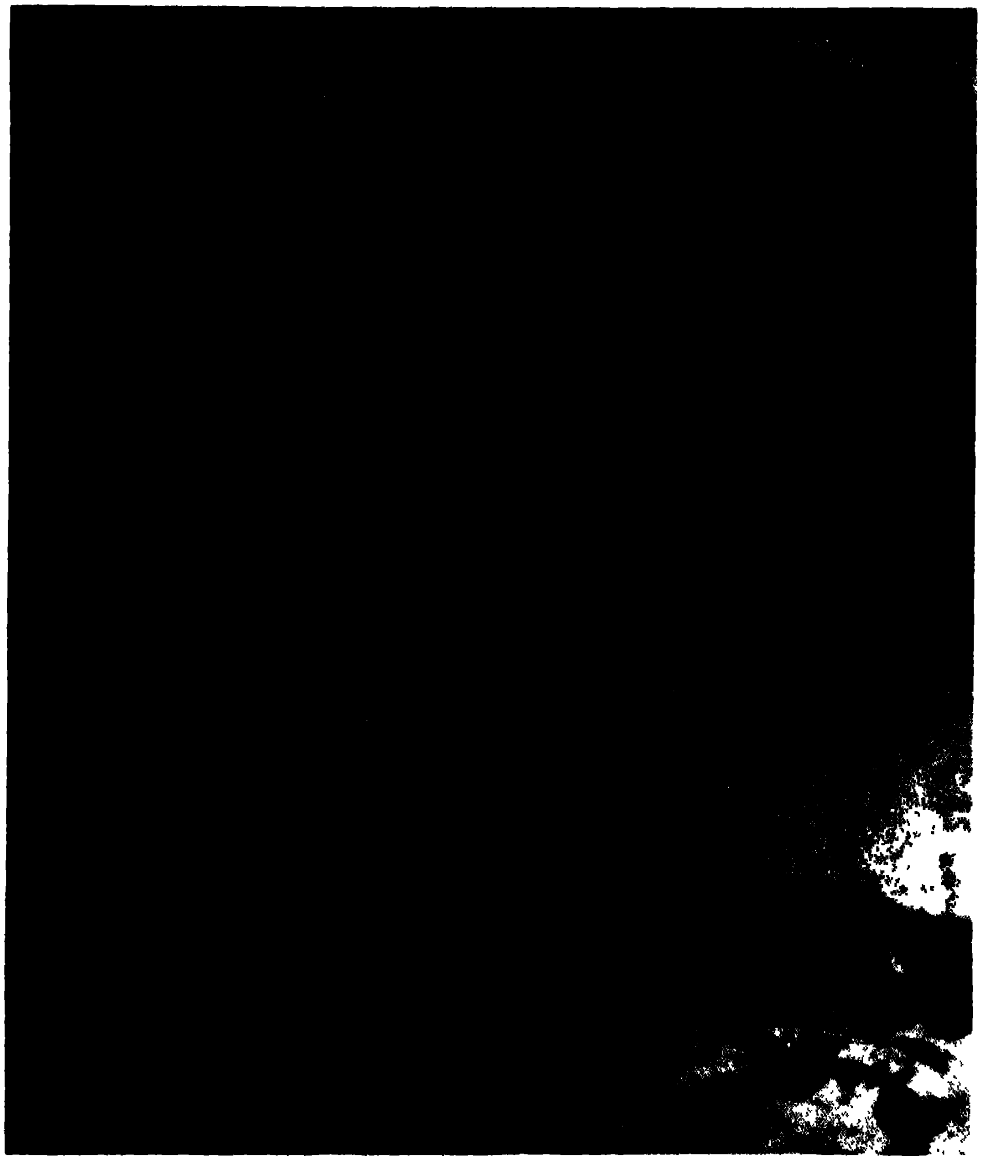


FIG. 7. Section of coronal dentine and pulp stained for acetylcholinesterase. Enzyme activity did not occur in the subodontoblast zone, the odontoblasts or in the dentinal tubules.

bundle. The third and most successful approach was to crack open the tooth and, with a sharp scalpel, cut round the periphery of the pulp chamber so as to include a millimetre or two of dentine. The pulp was carefully removed, taking with it this thin shell of dentine, and the whole was supported on a piece of liver, frozen and sectioned. The resulting sections contained coronal odontoblasts *in situ* in addition to predentine and some dentine. After staining these sections for acetylcholinesterase activity the enzyme was demonstrated in nerve fibres in the deeper part of the pulp. No enzyme activity occurred in the subodontoblast zone where histologically the nerve plexus of Rashkow exists, nor in the odontoblasts (Plate VIII, Fig. 7). One section contained dentine cut transversely which showed particularly well peritubular dentine and the odontoblast process *in situ* in the tubule. There was no acetylcholinesterase activity in the odontoblast process. Extended incubation in the substrate medium failed to produce a reaction in this site.

These findings indicate that acetylcholinesterase activity is present in the periodontal ligament, in the nerve trunks ascending in the pulp, and in fine non-myelinated fibres in the pulp. However, the findings of Avery and Rapp could not be repeated with respect to activity of this enzyme in the subodontoblast zone nor the odontoblast and its process, even though thinner sections and a more precise technique were used. This does not mean to say however, that the odontoblast is not implicated in dentine sensitivity but if it is, the impulse whatever its nature, is not dependent upon the activity of acetylcholinesterase.

REFERENCES

Avery, J. K. and Rapp, R. (1959) *J. dent. Res.*, **12**, 190.

Balogh, J. R. K. (1962) *J. Histochem.*, **10**, 232.

Brachet, J., Decrolly-Briers, M. and Hoyez, J. (1958) *Bull soc. Chim. Biol.*, **40**, 2039.

Diamond, M. and Weinmann, J. P. (1940) *The Enamel of Human Teeth.* Columbia University Press, New York.

Fearnhead, R. W. (1959) M.D.S. thesis. University of London.

Gomori (1941) *Arch. Path.*, **32**, 189.

Karnovsky, M. J. and Roots, L. (1964) *J. Histochem.*, **12**, 219.

KOELLE, G. B. and FRIEDENWALD, J. S. (1949) *Proc. Soc. exp. Biol., N.Y.*, **70**, 617.

McHUGH, W. D. (1961) *Dent. Practit.*, **11**, 314.

NOVIKOFF, A. B. (1961) *The Cell*, ed. Brachet, J. and Mirsky, A. F., 2, p. 423. Academic Press, New York.

OSBORN, J. (1965) Personal communication.

PEARSE, A. G. E. (1960) *Histochemistry. Theoretical and Applied* (2nd ed.). Churchill, London.

SYMONS, N. B. B. (1962) *Arch. oral Biol.*, **7**, 455.

TEN CATE, A. R. (1963) *Arch. oral Biol.*, **8**, 755.

TEN CATE, A. R. (1965) *Arch. oral Biol.*, **10**, 207.

X

Staphylococci Endemic in Hospitals

M. T. PARKER

Two separate but interconnected series of events account for the importance of certain strains of *Staphylococcus aureus* as causes of endemic sepsis in hospitals.

The first of these is the acquisition by patients—in the nose and elsewhere—of staphylococci not carried on admission. This is a very common event. As many as 10 per cent of patients may acquire a fresh strain of *Staph. aureus* during each week of stay in a hospital ward. Many patients are, of course, already carriers when they enter hospital, and the staphylococci they carry include a great variety of different strains, few of which subsequently spread to other patients. Most of the staphylococci acquired by patients in hospital, however, belong to a small number of strains endemic in the wards.

The second event is clinical sepsis, and this is much less common. Some of it, admittedly, is due to self-infection with staphylococci brought into hospital by the patients, but the majority is due to organisms first encountered after admission to hospital; and the bulk of hospital-acquired sepsis is due to organisms endemic in hospitals. Carriers form by far the most numerous sources of hospital staphylococci, and serve to maintain the continuity of infection with individual strains for long periods of time.

The strains of *Staph. aureus* now prevalent in hospitals differ in many ways from those commonly seen a few years ago. It is apparent, therefore, that evolutionary changes have taken place in staphylococci in the hospital environment, and recent investigations have shed light on the ways in which some of them have been brought about.

ANTIBIOTIC RESISTANCE AND THE SPREAD OF STAPHYLOCOCCI IN HOSPITALS

It was in the years after 1945, when an increasing proportion of *Staph. aureus* cultures from lesions in hospital patients were found to be resistant to penicillin (North and Christie, 1945; Roy and Greenberg, 1945), that attention was first drawn to the hospital as a source of endemic staphylococcal infection. A few strains of *Staph. aureus* isolated before 1940 were resistant to penicillin and formed penicillinase (Fleming, 1942; Knight and Holzer, 1954; Parker and Lapage, 1957), and it is possible that their ability to destroy penicillin may have served to protect them against antibiotic encountered in the natural environment, including that formed by dermatophytes (Smith and Marples, 1964). Barber (1947*a*, *b*) showed that penicillinase-producing staphylococci did not arise anew in each patient treated with penicillin, but were selected by the antibiotic and spread from patient to patient by cross-infection. The fact that many patients who were not treated with penicillin—as well as members of the hospital staff—also acquired resistant organisms was explained when Gould (1958) showed that antibiotic dust in the hospital environment also had a selective effect on the nasal flora. When penicillinase-producing staphylococci first became prevalent, nearly all had one of a very few phage-typing patterns (Barber and Whitehead, 1949), but as the years passed the variety of patterns among them increased, so that penicillinase-forming strains with nearly all possible phage-typing patterns can now be found.

The introduction of other antibiotics was followed by the appearance in hospitals of strains resistant to each of them; this happened successively with streptomycin, chloramphenicol, tetracycline, novobiocin, and neomycin, sometimes promptly, sometimes after a considerable interval. Staphylococci resistant to these antibiotics were not found among organisms isolated before the corresponding agent had been used. Their subsequent appearance is thought to have been due to the selection of naturally occurring mutants and their spread by cross-infection. Resistance to other antibiotics was found almost exclusively in penicillinase-producers. Once again, the first resistant organisms had one of a very few phage-typing patterns: but, unlike penicillinase-production, multiple antibiotic-resistance remained confined to a few strains, and there was a piling-up of resistance to more and more antibiotics in this minority of strains.

Between a quarter and a third of the staphylococci carried by people outside hospital are now resistant to penicillin (Miller, Galbraith and Green, 1962), but few of them are resistant to other antibiotics unless the carrier has had recent contact with a hospital (Galbraith, 1960). Staphylococci endemic in hospitals —with the exception of nurseries for the newborn—are nearly all resistant to two or more antibiotics, and in this country the second most common resistance is to tetracycline. In a recent Public Health Laboratory Service investigation of patients in medical wards, for example, the proportion of persons carrying penicillin-resistant *Staph. aureus* rose from 11 per cent on admission to 30 per cent in the fifth week and the proportion carrying tetracycline-resistant *Staph. aureus* rose from 4 per cent to 20 per cent. Stay in hospital resulted, therefore, in no net gain in the percentage of patients carrying organisms resistant only to penicillin. These organisms appear now to have been edged out of hospitals, and to be maintained in the general population.

A possible explanation of the restriction of multiple antibiotic-resistance to a minority of penicillin-resistant staphylococci comes from the work of Richmond, Parker, Jevons and John (1964) on the measurement of penicillinase-production. Fig. 1 shows the number of cultures producing different amounts of penicillinase under standard conditions. Strains resistant to two or more antibiotics usually had a high penicillinase activity, but those resistant only to penicillin produced on average much less. This suggested the following hypothesis. Staphylococci forming large amounts of penicillinase might be able to survive on carriage sites in hospital patients, and thus be subjected for long periods to the selective action of other antibiotics, while weaker penicillinase producers were eliminated. This view received some support from the observation that staphylococci isolated from rheumatic patients on prolonged penicillin prophylaxis formed on average greater amounts of penicillinase than those from patients not receiving penicillin (Michael, Michael and Massell, 1964).

The appearance in hospitals of methicillin-resistant strains of *Staph. aureus*—organisms partially insusceptible to the action of the 'penicillinase-resistant' penicillins—raises other evolutionary possibilities. These organisms existed in very small numbers

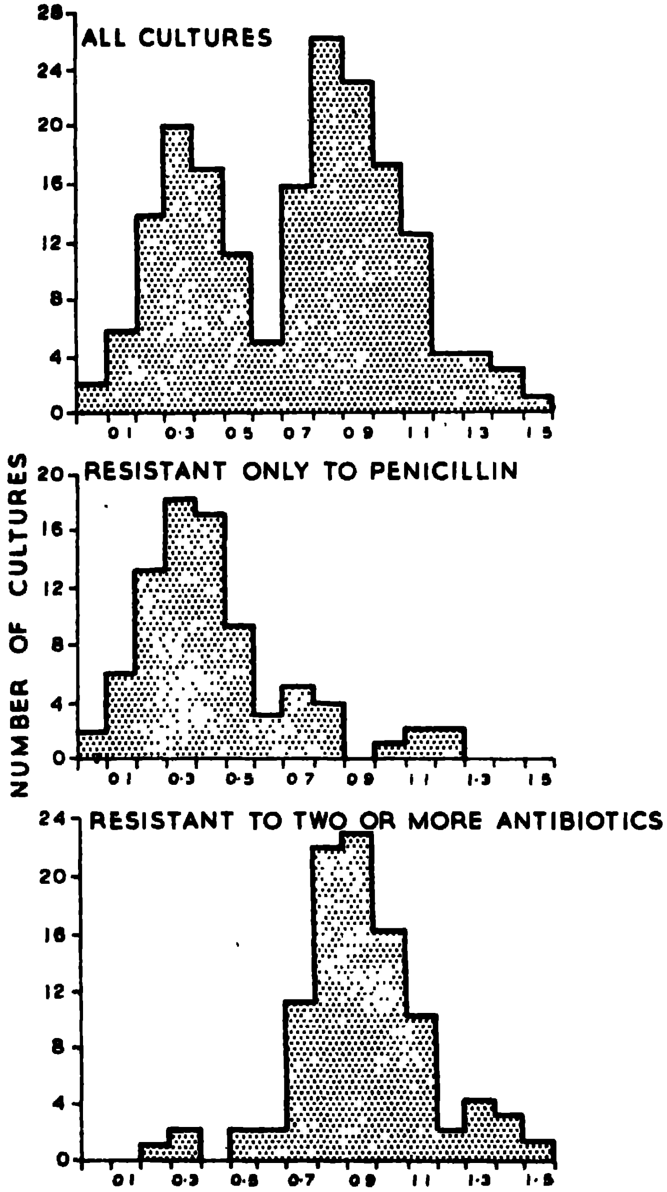


Fig. 1. Quantity of penicillinase synthesized by 181 cultures of *Staph. aureus* correlated with resistance to penicillin alone, and with multiple antibiotic resistance (Richmond *et al.*, 1964). A factor = 1·0 is equivalent to a penicillinase production of 330 enzyme units per mg bacteria.

before methicillin was used for treatment (Jevons, 1961), and have slowly become more common. In a large sample of *Staph. aureus* cultures isolated between 1960 and 1964, there was a rise in the proportion resistant from 0·05 per cent to 0·5 per cent. These organisms certainly produce large amounts of penicillinase, but their insusceptibility to methicillin is not due to this, because penicillinase-negative variants of them are resistant o methicillin and to other penicillins (Parker and Jevons, 1964). There is evidence suggesting that the increase in the prevalence of methicillin resistance may not have been due to the selective action of methicillin. For example, Borowski, Kamieńska and Rutecka (1964) in Poland, at a time when methicillin had not been used there, observed the colonization of half the babies in a nursery with a methicillin-resistant *Staph. aureus*. It seems possible, therefore, that methicillin resistance, which is really a general tolerance for all penicillins, may provide an auxiliary mechanism for survival in the presence of 'ordinary' penicillins, independent of penicillinase production and useful in environmental situations in which penicillinase is ineffective or is not produced in sufficient amount.

Virulence of Hospital Staphylococci

It is not easy to be sure that there really are differences between strains of *Staph. aureus* in their virulence for man. So far, laboratory tests have given little help, and we are dependent for our information on field investigations of the natural disease in human populations. There is, however, a good deal of evidence that strains do differ in their ability to cause septic lesions, based on differences in the frequency of particular phage-typing patterns between staphylococci from lesions and from nasal swabs of persons outside hospital (Williams and Jevons, 1961), and on observations that some strains spreading in particular wards give rise to lesions but others do not. (For example, see Barber, Wilson, Rippon and Williams, 1953; Shooter, Smith, Griffiths, Brown, Williams, Rippon and Jevons, 1958.) That *Staph. aureus* strains differ in their ability to cause severe and fatal infections is not so certain, because the infrequency of severe infections makes it difficult to exclude the effects of chance or of differences in host susceptibility.

The character of staphylococcal strains that has been most often measured is the ability to cause groups of lesions in a particular hospital or ward. This has been termed the *epidemic virulence* of a staphylococcus. Shooter and his colleagues (1958), in an often quoted investigation, showed that of 186 distinguishable strains of *Staph. aureus* in one ward, only thirteen caused disease, and only three were responsible for sepsis in more than one patient. The recognition of such strains is of practical importance, because infections with them are, theoretically at least, preventible by isolation (Williams, Noble, Jevons, Lidwell, Shooter, White, Thom and Taylor, 1962). But we must realize that we are measuring a composite character, which includes both the ability to become established in the ward, and more than a certain minimum of virulence for the average patient.

It is not surprising, therefore, to find that nearly all strains with epidemic virulence are resistant to several antibiotics. Indeed, tetracycline resistance is widely used in this country as a marker for epidemic virulence. Yet another marker, the significance of which was not at first apparent, is mercury resistance (Moore, 1960). Nearly all strains responsible for groups of lesions in hospital patients are resistant to the action of mercury salts. The clue to this association came when it was observed that mercury resistance was the character most closely correlated with the production of large amounts of penicillinase (Richmond *et al.*, 1964). This was taken further by Richmond and John (1964) who selected penicillinase-negative variants from mercury-resistant *Staph. aureus* cultures and found that they were mercury sensitive. Further, the properties of active penicillinase production and mercury resistance could be co-transduced to penicillinase-negative variants, both of originally mercury-resistant, and of originally mercury-sensitive staphylococci (Table 1). The fact that a small proportion of the transductants had only one of the two characters indicated that the association was not a functional one, but was due to a close linkage of the gene responsible for mercury resistance with a gene responsible for high penicillinase activity.

So, a test for mercury resistance in *Staph. aureus* is effectively a test for the ability to produce large amounts of penicillinase; and these characters have probably come to be associated by selection in the hospital with multiple antibiotic-resistance and tetracycline resistance. But the markers for epidemic virulence are in reality only 'one-way' markers. A strain with the ability to cause groups of lesions usually possesses them, but their presence in a strain does not guarantee that it has epidemic virulence. Their absence, however, suggests that a strain is unlikely to spread widely enough in a hospital to cause many lesions, however virulent it may be for the individual patient.

TABLE 1. Co-transduction of mercury resistance and penicillinase production (Richmond and John, 1964)

Transducing phage was prepared from a penicillinase-producing, mercury-resistant (P/R.Hg/R) parent.
The acceptor strains were penicillinase-negative (P/S) variants from (1) a P/R.Hg/R and (2) a P/R.Hg/S culture, and were both mercury-sensitive (Hg/S).
In experiment A the transductants were selected for P/R, and then tested for Hg/R; in experiment B they were selected for Hg/R and then tested for P/R.

Acceptor		Transductants	
Original	Phenotype		
	A. Selection for P/R		
		P/R.Hg/R	P/R.Hg/S
P/R.Hg/R	P/S.Hg/S	302	2
P/R.Hg/S	P/S.Hg/S	73	1
	B. Selection for Hg/R		
		P/R.Hg/R	P/S.Hg/R
P/R.Hg/R	P/S.Hg/S	146	2
P/R.Hg/S	P/S.Hg/S	472	7

THE CHANGING PATTERN OF ENDEMIC PREVALENCE IN HOSPITALS

We shall now consider the history of some of the common hospital strains that have the character of epidemic virulence, making use of information collected in the Staphylococcus Reference Laboratory during the last ten years (Williams, 1959; Williams and Jevons, 1961; Parker and Jevons, 1963) from cultures sent to us by pathologists seeking help in the control of hospital sepsis. We adopted the convention that three cultures with the same phage-typing pattern from lesions in one hospital during one year was evidence of an 'epidemic', though perhaps it would have been more accurate to describe it as an endemic prevalence. Fig. 2 shows the percentage of such 'epidemics' that were due to organisms with a particular phage-typing pattern, and it is clear that, at any one time, over half of them were due to a relatively small number of organisms, as defined by phage-typing. In the years 1954–7, for example, four strains, with the patterns 80 (or, as we should say now, 80/81), 52A/79, 75/77, and 47/53/75/77, were together responsible for 51 per cent of the 'epidemics'. In later years, however, there were considerable changes in the relative importance of the more common strains. By 1962–3, the 80 or 80/81 staphylococci, though still common, had been overshadowed by related members of the 52, 52A, 80, 81 complex; and two other sorts of organism, those lysed by phage 83A and a related group until recently untypable ('new') had become prominent.

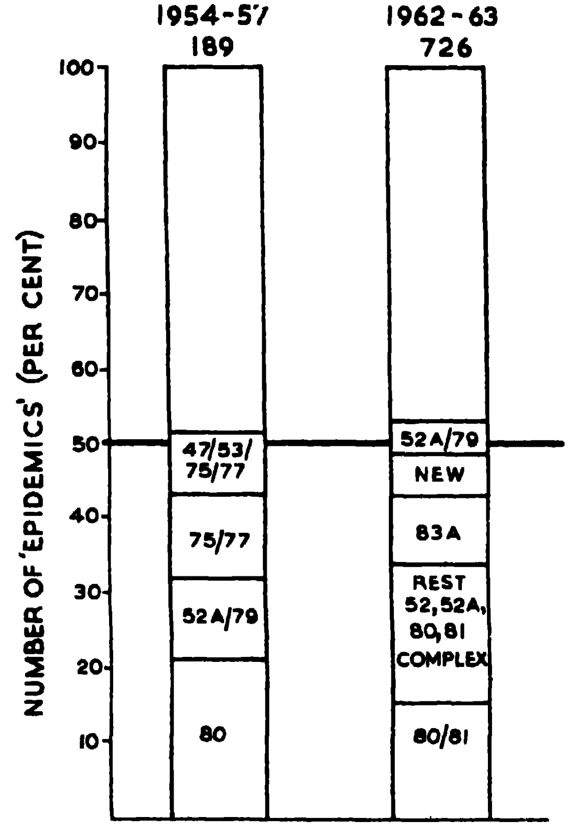


FIG. 2. 'Epidemics' of sepsis in hospitals, 1954–57 and 1962–63. Percentage due to *Staph. aureus* strains with certain phage-typing patterns. Phages 81 and 83A were not used in 1954–57. The patterns given are the modal typing patterns at routine test dilution. New = untypable strains with inhibition patterns similar to those of 83A staphylococci.

It would be wrong, however, to conclude that these few strains were the only ones of significance as causes of outbreaks of sepsis. Their predominance was due, in part at least, to the fact that they had established themselves in a large number of different hospitals. Plenty of other strains gave rise to outbreaks, sometimes with severe consequences for the patients; it was simply that they did so less often.

THE 52, 52A, 80, 81 COMPLEX

We have learned a good deal about the evolution of hospital staphylococci from the events which followed the appearance of the 80/81 organism as a cause of sepsis in hospitals. It was probably recognized in many countries soon after it became prevalent, because the outbreaks to which it gave rise were unusually severe. The first of these, in Australia in mid-1953 (Isbister, Durie, Rountree and Freeman, 1954; Rountree and Freeman, 1955), was due to an organism that was an active producer of penicillinase and was mercury resistant, but was sensitive to other antibiotics. Similar organisms—but resistant also to tetracycline—were isolated a few months later in Canada, and within a year in the United States and elsewhere (Bynoe, Elder and Comtois, 1956. See also Rountree and Beard, 1958). It is difficult to be certain that it was the same organism which had spread around the world so rapidly, but the state of lysogeny of cultures isolated from widely separated places is so characteristic as to suggest that this may well have been so.

The 80/81 organism was first identified in England in 1954, and had become the most common single cause of outbreaks of sepsis by 1955 (Fig. 3). Then, in 1957, strains with the typing pattern 52/52A/80 (or 52/52A/80/81 when phage 81 was introduced in 1961) appeared, particularly in hospitals in which the 80/81 strain had been endemic. Staphylococci with the typing pattern 52/52A/80/81, but penicillin sensitive, had caused sporadic infections outside hospital for many years. It was soon found that strains with this typing pattern carried phages which could change the typing pattern of an 80/81 organism to 52/52A/80/81 by lysogenization (Rountree, 1959; Asheshov and Rippon, 1959). It seemed likely, therefore, that when

52/52A/80/81 strains appeared in hospitals, they had arisen from the 80/81 organism by the acquisition of phages from native 52/52A/80/81 strains.

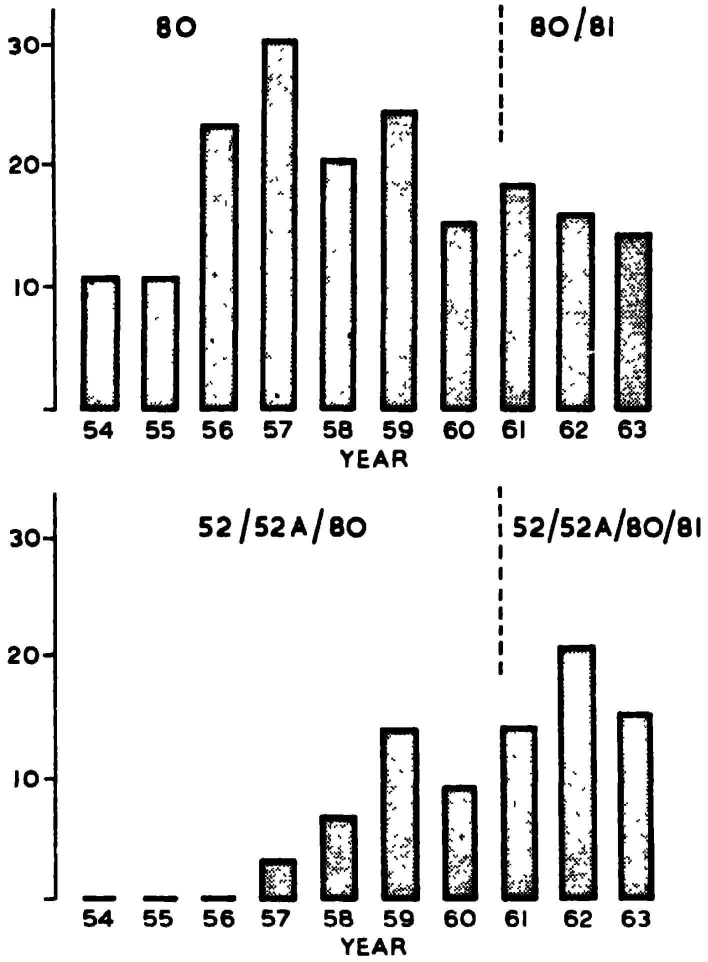


Fig. 3. Percentage of 'epidemics' of hospital sepsis due to strains of *Staph. aureus* with the phage-typing patterns 80 (or 80/81) and 52/52A/80 (or 52/52A/80/81). Phage 81 was first used in 1961.

It was surprising, however, that the acquisition of a phage should have widened the phage-typing pattern. The explanation was provided by Rountree and Asheshov (1961), who showed that the insusceptibility of 80/81 strains to phages 52 and 52A was due to the carriage of another phage—a completely defective phage we call 80′—without which its phage-typing pattern is 52/52A/80/81 (Fig. 4). When 80/81 strains are lysogenized with certain phages (Type 1 converting phages) the 80′ phage is lost by prophage substitution, and the typing pattern becomes 52/52A/80/81. Other phages (Type 2) both substitute for 80′ and block the 81 reaction, giving the pattern 52/52A/80. Some phages block one or more of the reactions with the typing phages without prophage substitution, giving the patterns 80 or 81, or producing an untypable organism. All these varieties occur naturally and carry the expected phages.

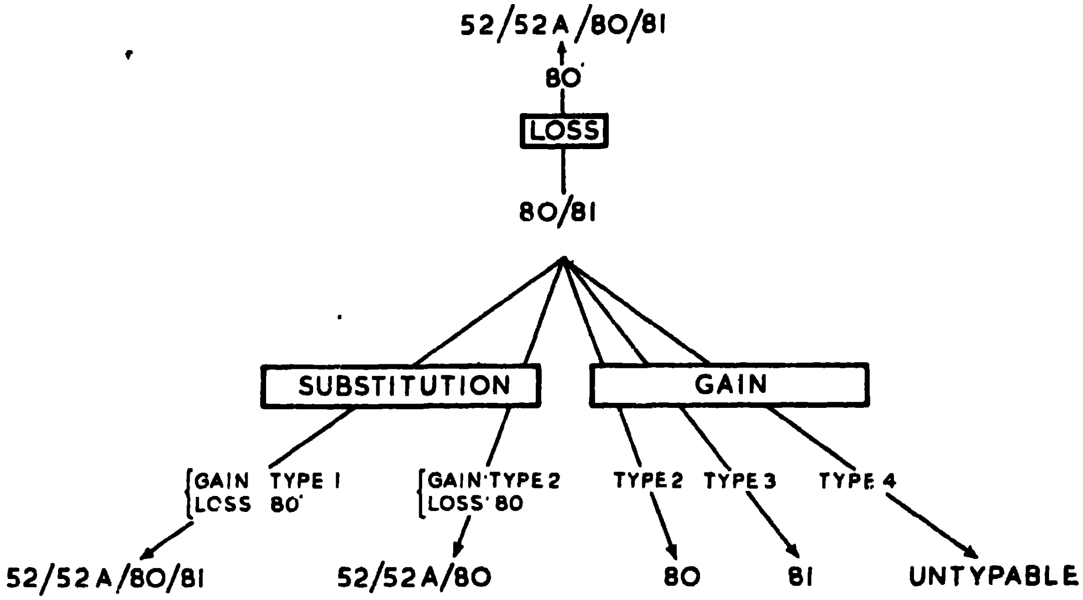


FIG. 4. The 52, 52A, 80, 81 complex. Relation of phage-typing pattern to lysogeny. 80′: defective phage carried by 80/81 strains. Types 1–4: converting phages which change the typing pattern of 80/81 strains on lysogenization.

It seems, therefore, that the members of the 52, 52A, 80, 81 complex have arisen from the original 80/81 organism as a result of a number of separate lysogenizations, but have retained the epidemic virulence of their ancestor. The differences in phage-typing pattern conceal an underlying similarity between the organisms.

A NEW UNTYPABLE ORGANISM RELATED TO THE 83A STAPHYLOCOCCUS

The next happening we shall consider was one in which a relatively simple and apparently uniform change in phage-typing pattern was associated with a complex series of alterations in the other characters of the organism.

In 1958 we had become aware that a previously unrecognized strain of *Staph. aureus* was causing outbreaks of sepsis in this country. It was not typable with the basic set of typing phages, but could be recognized by a series of inhibition patterns with

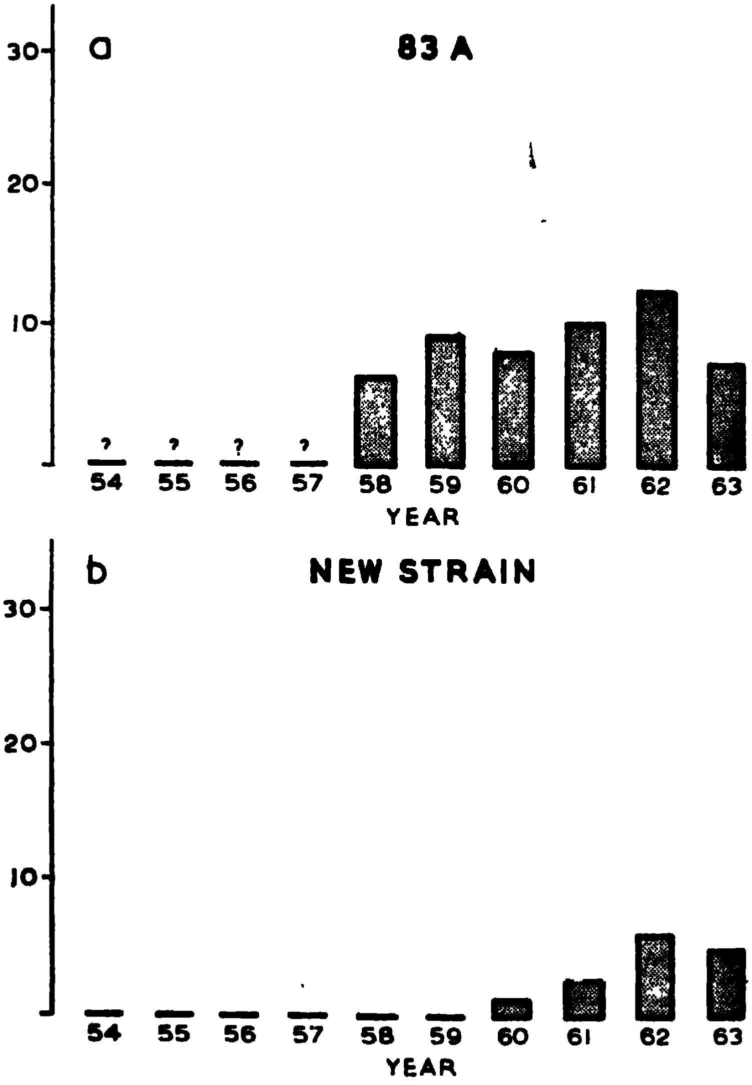


FIG. 5. Percentage of 'epidemics' of hospital sepsis due to *Staph. aureus* (*a*) lysed only by phage 83A, (*b*) untypable strains with inhibition patterns similar to those of 83A staphylococci (new untypable staphylococci).

group III phages applied at 1,000 times the routine test dilution. It was soon found that this organism could be characterized by a lytic reaction with an experimental phage (83A), and this phage was later introduced into the basic typing set. We know nothing about the origin of this 83A organism, and are not sure exactly when it first appeared, but Fig. 5 shows the percentage of 'epidemics' it caused from 1958 onwards.

Two years later, untypable organisms with the same inhibition reactions, but not lysed by phage 83A, began to appear. Later they became more common (Fig. 5) and gave rise to extensive outbreaks of sepsis in many hospitals. It seemed possible that this new untypable organism might have arisen from the 83A strain by the acquisition of a phage which blocked the 83A reaction. Several strains of the new organism were examined, and each was found to carry a phage which, when used to lysogenize an 83A staphylococcus, rendered it insusceptible to lysis by phage 83A; and non-lysogenic variants of the new untypable organism had a phage-typing pattern indistinguishable from that of the 83A organism (Jevons and Parker, 1964).

Examination of the blocking phages from several distinct strains of the new untypable organism showed that they were all different, either in serological group or host-range, or in both. Further evidence of the heterogeneity of the new organism came from our attempts to find a phage to characterize it by a lytic reaction. It was easy enough to find phages which lysed some of the new organisms, but none lysed them all. So, with a set of five such phages, we obtained many different pattern-reactions. When we tried to reproduce the whole range of these pattern-reactions by lysogenizing 83A organisms, we found it necessary not only to use several different blocking phages, but to start from more than one different 83A organism (Jevons and Parker, 1964). It appeared, therefore, that the new untypable organism had arisen by a number of separate lysogenizations, and probably not all from the same ancestor.

We were surprised, therefore, when reports appeared from different parts of the country that many of the new untypable organisms possessed a series of unusual cultural characters. These included the formation of a characteristic lemon-yellow pigment on glycerol monoacetate agar, the production of β-toxin and the absence of fibrinolysin (staphylokinase), and neomycin resistance (Willis and Turner, 1963; Willis, Jacobs and Goodburn, 1963; Robertson, 1963). Two obvious explanations, that these were simply the characters of a locally prevalent strain, or were characters shared by the new organism with the original 83A organism, were soon disposed of. We examined a collection of twenty-two 83A cultures, and twenty-seven representatives of the new untypable organism, chosen to include distinct strains from many different parts of Great Britain. The two sets of cultures had several characters in common, including resistance to penicillin, streptomycin, tetracycline, and mercury salts, and usually also to erythromycin, but there were five distinct differences between them (Fig. 6). All of the 83A cultures were sensitive to neomycin, and over 90 per cent hydrolysed Tween 80; few formed the lemon-yellow pigment; over half were fibrinolysin positive but very few formed β-toxin. On the other hand, three-quarters of the new untypable organisms were neomycin resistant, and less than a quarter split Tween 80; three-quarters were lemon-yellow; less than 10 per cent were fibrinolysin positive but 40 per cent formed β-toxin.

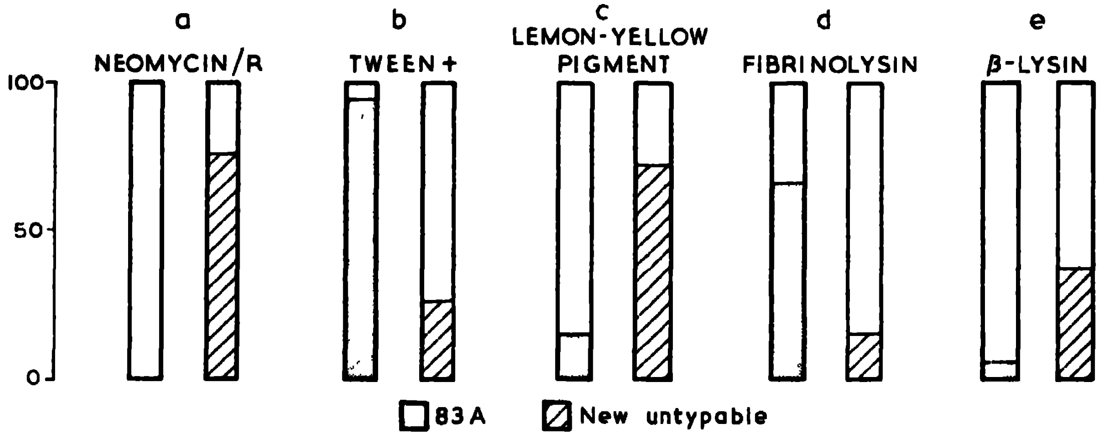


FIG. 6. Cultural characters of 83A staphylococci (22 cultures) and of new untypable staphylococci (27 cultures). Percentage of each (*a*) resistant to neomycin (5 μg/ml), (*b*) splitting Tween 80, (*c*) giving lemon-yellow pigmentation on glycerol monoacetate agar, (*d*) producing fibrinolysin, and (*e*) β-toxin positive.

These differences might be explained if it could be shown that some of the characters were linked, and were related to the carriage of the phages which had resulted in the change of phage-typing pattern. This seemed possible for the following reasons. Rosendal, Bülow and Jessen (1964) had shown that the acquisition by 80/81 staphylococci of a phage which blocked the 81 reaction also changed the Tween reaction from positive to negative. This happened in every lysogenized cell and was therefore due to lysogenic conversion. It has also been known for many years that *Staph. aureus* cultures were seldom both β-toxin and fibrinolysin positive (Christie and Wilson, 1941). The reason for this was apparent when Winkler and his colleagues (de Waart, Winkler and Grootsen, 1962; Winkler, de Waart and Grootsen, 1965) found that the acquisition of group F phages resulted in many *Staph. aureus* strains in the change from β-toxin positive, fibrinolysin negative to β-toxin negative, fibrinolysin positive, and that this was also due to lysogenic conversion.

We found that the change from Tween positive to Tween negative, and the acquisition of lemon-yellow pigmentation could both be brought about by lysogenization of 83A strains with phages which blocked the 83A reaction. Some of the blocking phages changed the Tween reaction, some changed the pigmentation, and some changed both; but lemon-yellow pigmentation could be detected after lysogenization only if the recipient 83A culture did not form a strong orange pigment. On the other hand, the presence or absence of these prophages had no effect on the production of β-toxin or fibrinolysin. Therefore, though the appearance of the characters β-toxin positive, fibrinolysin negative in the new untypable organism might have been due to the loss of a carried phage, this loss was not brought about by prophage-substitution when the blocking phage was acquired. Finally, loss or gain of blocking phages had no effect on the sensitivity of the organisms to neomycin.

If several separate events had to take place to produce a new untypable organism with its typical pattern of cultural characters, some naturally occurring strains should show deviations from this pattern. This was found to be so. We examined five sets of new untypable organisms each from a different London hospital in which the organism was endemic. In three of the hospitals, the predominant organism had the expected characters (neomycin resistant, Tween negative, lemon yellow, β-toxin positive, fibrinolysin negative). The organism endemic in one hospital was β-toxin negative, fibrinolysin positive, but otherwise typical. In the fifth hospital, the predominant organism was neomycin sensitive, orange-pigmented, β-toxin negative and fibrinolysin positive, but Tween negative.

It might be thought that the neomycin resistance of the new untypable organism was entirely responsible for its recent success as a hospital staphylococcus, but some facts remain to be explained. Neomycin resistance, though rare among other *Staph. aureus* strains, is not unknown. Examination of a large collection of 83A strains in 1964 revealed that nearly one per cent were neomycin resistant. This might suggest that the predominant neomycin-resistant untypable organism had arisen from one of the few neomycin-resistant 83A strains. If so, it is difficult to understand why the neomycin-resistant 83A organism did not itself become prevalent. While not denying the selective advantage conferred by neomycin resistance in the hospital environment, the observed sequence of events suggests that, in the new untypable organism, it came to be associated with some other advantageous character we have not yet identified.

It is clear that we are a long way from understanding what are the essential characters of successful hospital staphylococci, but we do have an inkling of the sort of evolutionary processes which have led to their emergence. We have known for many years that conditions in hospitals favoured the spread of certain strains of *Staph. aureus* once they had arisen. It now appears that the situation in hospitals is also one in which strains with new combinations of characters arise frequently. The hospital is a sort of cross-roads for staphylococci, a meeting-place in which there are almost unlimited possibilities for the loss or gain of genetic material, either by prophage exchange or, for all we know, by other mechanisms as well.

REFERENCES

ASHESHOV, E. H. and RIPPON, J. E. (1959) *J. gen. Microbiol.*, **20**, 634.
BARBER, M. (1947*a*) *J. Path. Bact.*, **59**, 373.
BARBER, M. (1947*b*) *Brit. med. J.*, **ii**, 863.
BARBER, M. and WHITEHEAD, J. E. M. (1949) *Brit. med. J.*, **ii**, 565.
BARBER, M., WILSON, B. D. R., RIPPON, J. E. and WILLIAMS, R. E. O. (1953) *J. Obstet. & Gynaec.*, **60**, 476.
BOROWSKI, J., KAMIEŃSKA, K. and RUTECKA, I. (1964) *Brit. med. J.*, **i**, 983.
BYNOE, E. T., ELDER, R. H. and COMTOIS, R. D. (1956) *Canad. J. Microbiol.*, **2**, 346.

CHRISTIE, R. and WILSON, H. (1941) *Aust. J. exp. Biol. med. Sci.*, **19**, 329.

FLEMING, A. (1942) *Lancet*, **i**, 732.

GALBRAITH, N. S. (1960) *Proc. R. Soc. Med.*, **53**, 253.

GOULD, J. C. (1958) *Lancet*, **i**, 489.

ISBISTER, C., DURIE, E. B., ROUNTREE, P. M. and FREEMAN, B. M. (1954) *Med. J. Aust.*, **ii**, 897.

JEVONS, M. P. (1961) *Brit. med. J.*, **i**, 124.

JEVONS, M. P. and PARKER, M. T. (1964) *J. clin. Path.*, **17**, 243.

KNIGHT, V. and HOLZER, A. R. (1954) *J. clin. Invest.*, **33**, 1196.

MICHAEL, T. M., MICHAEL, J. G. and MASSELL, B. F. (1964) *Amer. J. med. Sci.*, **248**, 152.

MILLER, D. L., GALBRAITH, N. S. and GREEN, S. (1962) *Brit. J. prev. soc. Med.*, **16**, 203.

MOORE, B. (1960) *Lancet*, **ii**, 453.

NORTH, E. A. and CHRISTIE, R. (1945) *Med. J. Aust.*, **ii**, 44.

PARKER, M. T. and JEVONS, M. P. (1963) In *Infection in hospitals*, ed. R. E. O. Williams and R. A. Shooter, Oxford, Blackwell, p. 55.

PARKER, M. T. and JEVONS, M. P. (1964) *Postgrad. med. J.*, **40**, Dec. (Suppl.), 170.

PARKER, M. T. and LAPAGE, S. P. (1957) *J. clin. Path.*, **10**, 313.

RICHMOND, M. H. and JOHN, M. (1964) *Nature, Lond.*, **202**, 1360.

RICHMOND, M. H., PARKER, M. T., JEVONS, M. P. and JOHN, M. (1964) *Lancet*, **i**, 293.

ROBERTSON, J. J. (1963) *Lancet*, **ii**, 333.

ROSENDAL, K., BÜLOW, P. and JESSEN, O. (1964) *Nature, Lond.*, **204**, 1222.

ROUNTREE, P. M. (1959) *J. gen. Microbiol.*, **20**, 620.

ROUNTREE, P. M. and ASHESHOV, E. H. (1961) *J. gen. Microbiol.*, **26**, 111.

ROUNTREE, P. M. and BEARD, M. A. (1958) *Med. J. Aust.*, **ii**, 789.

ROUNTREE, P. M. and FREEMAN, B. M. (1955) *Med. J. Aust.*, **ii**, 157.

ROY, T. E. and GREENBERG, L. (1945) *J. Canad. med. Serv.*, **2**, 116.

SHOOTER, R. A., SMITH, M. A., GRIFFITHS, J. D., BROWN, M. E. A., WILLIAMS, R. E. O., RIPPON, J. E. and JEVONS, M. P. (1958) *Brit. med. J.*, **i**, 607.

SMITH, J. M. B. and MARPLES, M. J. (1964) *Nature, Lond.*, **201**, 884.

WAART, J. DE, WINKLER, K. C. and GROOTSEN, C. (1962) *Nature, Lond.*, **195**, 407.

WILLIAMS, R. E. O. (1959) *Lancet*, **i**, 190.

WILLIAMS, R. E. O. and JEVONS, M. P. (1961) *Zbl. Bakt., I. Abt. Orig.*, **181**, 349.

WILLIAMS, R. E. O., NOBLE, W. C., JEVONS, M. P., LIDWELL, O. M., SHOOTER, R. A., WHITE, R. G., THOM, B. T. and TAYLOR, G. W. (1962) *Brit. med. J.*, **ii**, 275.

WILLIS, A. T., JACOBS, S. I. and GOODBURN, G. M. (1963) *Lancet*, **ii**, 67.

WILLIS, A. T. and TURNER, G. C. (1963) *J. Path. Bact.*, **85**, 395.

WINKLER, K. C., DE WAART, J. and GROOTSEN, C. (1965) *J. gen. Microbiol.*, **39**, 321.

XI

Antiviral Chemotherapy

D. J. BAUER

THERE are, as yet no antiviral drugs in the British Pharmacopoeia, but this does not mean to say that there will not be in the comparatively near future. There are already two antiviral drugs which are being extensively used in man, and it will be the object of this review to try to demonstrate that not only does antiviral chemotherapy exist, but that it is being put into practice on an ever-increasing scale, and that we more or less stand in the same position with regard to the treatment of virus infections as we did with bacterial diseases just after the introduction of the sulphonamides.

The literature on antiviral chemotherapy abounds with compounds which appear to have antiviral activity *in vitro*, but very few of these have any effect when tested against virus infections in experimental animals. Of these, only three have been used extensively in man, and one has not fulfilled its earlier promise. It is not my intention to review this enormous mass of work which has led to so little practical result, but rather to put forward the principles from experience accumulated in the development of an antiviral drug with which I have been concerned for many years.

The first, and probably the most important principle relating to the field of antiviral chemotherapy is to be personally convinced that the discovery of antiviral drugs is a practical possibility. There has been much defeatism in this respect. Antibacterial chemotherapy has been in existence for 25 years, and has made enormous advances, each of which has seemed to underline the difficulty, or even the impossibility, of finding effective antiviral agents. However, the advances of bacterial chemotherapy have been largely due to luck, luck in the sense that moulds will readily produce complicated and highly effective molecules which no chemist would ever have considered synthesizing. The lack of success in the antiviral field can be ascribed to the unfortunate fact that these moulds do not oblige us by synthesizing the right compounds, coupled to the self-evident fact that those compounds which we have synthesized ourselves are in most cases all too clearly the wrong ones.

The second principle in antiviral chemotherapy, then, is to have some means of picking out the right compounds. This is rather like filling up a football coupon. One can have a system, or one can make a selection at random. Let us take the second method first, if only to dismiss it. Many thousands of compounds chosen at random have been examined for antiviral activity, without any result of practical importance.

We are therefore thrown back upon system. Whether this yields any results depends of course upon whether one has selected the right system. In my opinion the most effective system is to try against virus infections any class of compounds which shows interesting activity in some other biological field, usually a field in which results are easier to come by. Methisazone, a drug effective in the prophylaxis of smallpox, was discovered as a result of the application of this system. The original discovery dates from 1946, when Domagk, Behnisch, Mietsch, and Schmidt prepared a number of derivatives of benzaldehyde thiosemicarbazone and found that they were active against tuberculosis. What prompted Domagk to try thiosemicarbazones I don't know, but he must have had a good reason for it as he had an uncanny facility for picking winners. I always think it a pity that he did not have a go at antiviral chemotherapy, but at least that means that he has left something for the rest of us to find out.

However, the possibilities of his work were not lost on the virologists, who were quick to try this new class of compounds against their particular viruses. Thus, Hamre, Bernstein and Donovick (1950) found that benzaldehyde thiosemicarbazone was active against vaccinia virus infection in mice and fertile eggs.

This was the first genuine antiviral agent, and vaccinia was the first true virus to succumb to antiviral chemotherapy. In retrospect, it is chastening to think that this knowledge was available in 1950, but only came to practical fruition in 1962. It is even more chastening to reflect that benzaldehyde thiosemicarbazone was first synthesized in 1904, and had been lying unappreciated in the chemical literature ever since. If its antiviral activity had been discovered in 1904 I am quite sure that antiviral chemotherapy would by now have progressed to unimaginable achievements, and that antibacterial chemotherapy, in spite of all the means which we have available now, would appear to be lagging far behind in comparison. My regret as a virologist at what might have been is somewhat tempered by the parallel fact that the bacteriologists have not done all that much better either, since sulphanilamide was first synthesized in 1908 and its potentialities were only discovered thirty years later.

The discovery of the activity of benzaldehyde thiosemicarbazone against vaccinia virus led to the examination of other thiosemicarbazones, and interest soon became centred upon isatin 3-thiosemicarbazone. This was among a number of thiosemicarbazones found active by Thompson, Minton, Officer and Hitchings (1953) who, however, did not develop their observations very far. Looking back now, it seems strange that their work did not arouse greater interest, but the climate of opinion at that time was not favourable for devoting resources to the discovery of antiviral agents.

I became interested in isatin 3-thiosemicarbazone, and I soon found that its antiviral activity was much higher than had been found by Thompson and his co-workers (Bauer, 1955). It would in fact protect mice which had been infected intracerebrally with 100,000 LD50 of neurotropic vaccinia virus. Activity of such a high order was obviously of clinical interest. Moreover, in these experiments the animals had been infected first and treated afterwards, during the incubation period. The effect which was observed was therefore prophylactic, and in view of the close relationship between vaccinia and smallpox viruses, it seemed to me that the compound might be useful for the prevention of smallpox in persons who had already been exposed to the infection. I was able to put this to the test eight years later, with successful results, but before this became a practical possibility a great deal of work remained to be done.

A typical experiment by which the antiviral activity of isatin 3-thiosemicarbazone can be demonstrated is shown in Table 1. The experiment also shows in a crude way the effect of making alterations in the structure of a molecule.

TABLE 1. Effect of isatin 3-thiosemicarbazone and derivatives in mice infected intracerebrally with vaccinia virus

Compound	Dose		Dilution of virus					
	mg	mg/kg	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
Isatin 3-thiosemicarbazone	2	100	4/6	1/6	0/6	0/6	0/6	0/6
1-Methylisatin 3-thiosemicarbazone	0·5	25	0/6	0/6	0/6	0/6	0/6	0/6
Isatin 3-semicarbazone	2	100	6/6	6/6	6/6	6/6	5/6	4/6
Control	—	—	6/6	6/6	6/6	6/6	5/6	4/6

Four groups of mice were inoculated intracerebrally with serial decimal dilutions of neurovaccinia virus, six mice being used for each dilution in each group. The animals of the first group were dosed with isatin 3-thiosemicarbazone in a dose of 2 mgm given subcutaneously. The first dose was given a few hours after infection, and dosing was continued twice a day for 5 days. The second group was given 1-methylisatin 3-thiosemicarbazone according to the same schedule. The third group was similarly given isatin 3-semicarbazone, and the fourth group was left untreated as a control.

Several things appear from the results. Firstly, the very high degree of activity exerted by isatin and 1-methylisatin 3-thiosemicarbazones, as shown by the absence of deaths in the treated animals as compared with the 100 per cent mortality in the controls up to a dilution of 10^{-4}. The compounds thus conferred absolute protection against more than 100,000 LD50 of vaccinia virus.

The second thing to be noted is that the substitution of a methyl group in the 1-position appears to increase activity. A defect of this particular test method now becomes apparent, since it is impossible by means of it to obtain a sufficiently accurate quantitative estimation of the effect of this particular substitution, and this point I will refer to below.

The third thing to be noticed is the complete lack of activity of isatin 3-semicarbazone. It is evident that replacement of the sulphur atom in the side-chain by oxygen completely abolishes the antiviral activity.

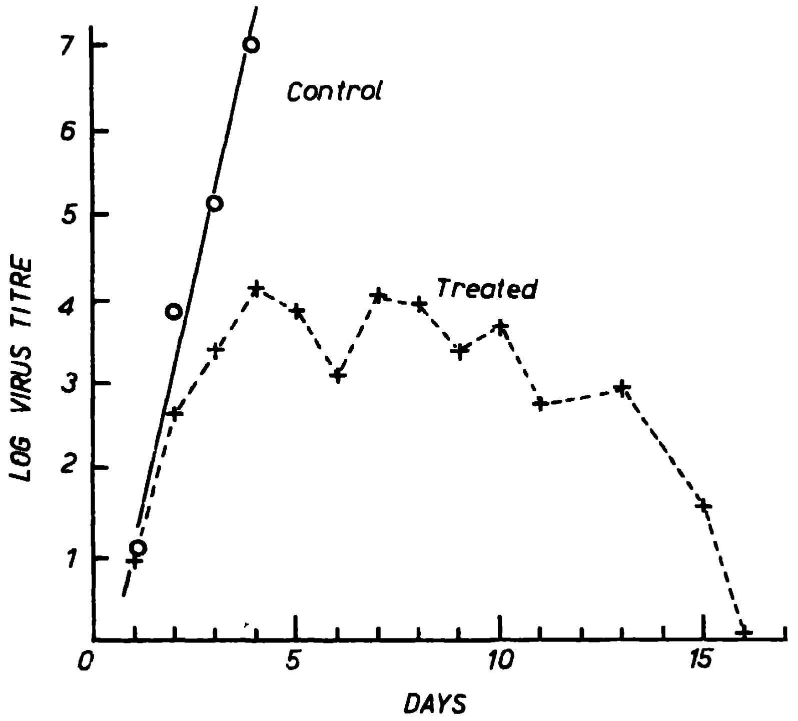


FIG. 1. Effect of treatment with methisazone upon the multiplication of vaccinia virus in mouse brain.

Of course, the fact that a compound protects against death does not mean that it is affecting the multiplication of the virus. In order to avoid false positive results it is therefore necessary to investigate whether the rate of multiplication of the virus is actually affected by the treatment which we have been applying.

In an experiment illustrated in Fig. 1 the continuous line represents the amount of vaccinia virus present in the brains of the untreated control mice on successive days after infection. The ordinate is logarithmic and by the fourth day the virus has multiplied until there is 100,000 times as much as was present initially. At this point all the mice died and the curve ends. The dotted line shows the growth of virus in the brains of mice infected in a similar manner but treated with 1-methylisatin 3-thiosemicarbazone. The compound was given in a dose of 2 mgm twice a day for a total of nine doses, after which treatment was discontinued. By the second day the line lies substantially beneath that of the controls and its gradient is declining, indicating that the growth of the virus is being slowed up. By the fourth day the growth of the virus has been brought to a halt. At this stage the treatment was discontinued, but no further growth of virus occurred and it gradually disappeared from the brain over a period of 20 days.

We therefore have here a true antiviral agent of high potency, and I should like to mention, in passing, that unless a substance can actually arrest the growth of virus in experimental animals in this sort of way it is not good enough to be considered for potential clinical use.

We now come to the third principle of antiviral chemotherapy. We have seen that modifying the molecule will alter the antiviral activity, and that one particular alteration, substituting a methyl group in the 1-position, appears to increase it. This holds out hope that other alterations might increase the activity still more, but before work of this kind can be put on a firm footing it is necessary to devise an assay method which will enable the activity of a derivative to be evaluated numerically in comparison with a standard.

Today this sounds rather like a statement of the obvious, but in 1956, when I began to work on the problem, there were no antiviral assay procedures in existence, for the simple reason that there were to all intents and purposes no antiviral compounds.

The pharmacologist faced with the problem of measuring the activity of a drug must first of all devise an experiment which will enable him to obtain a dose-response curve. In the test which I have already described the response is quantal, the mice either die or survive, and for a dose-response curve it is much better to find a response which is a continuous variable. I therefore investigated the effect of giving the infected animals graded doses of compound which were too small to confer protection against death, in order to see whether the survival time was related in any way to the dose of compound given.

An experiment of this type is shown in the Fig. 2. Six groups of six mice each were all given 1,000 LD50 of vaccinia virus intracerebrally. The first and last groups were left untreated as controls, and the other four groups were treated with 7-ethylisatin 3-thiosemicarbazone in doses of 0·005, 0·01, 0·02 and 0·04 mgm respectively. The first dose was given some hours after infection and dosing was continued twice a day for five days. The time of death of the animals was recorded to the nearest half-day, and animals which showed no signs of illness by the fourteenth day were discarded and considered to have survived indefinitely.

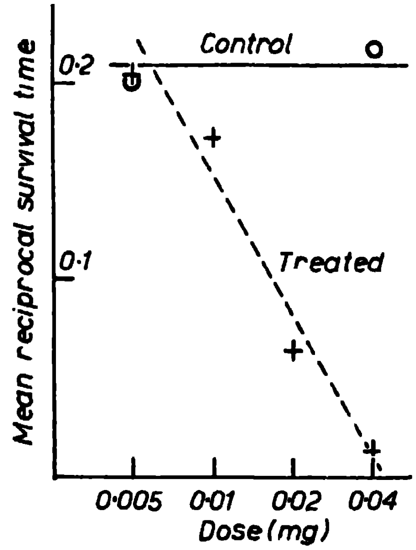


FIG. 2. Dose-response line of 7-ethylisatin 3-thiosemicarbazone in mice infected intracerebrally with vaccinia virus.

Some relationship between dose of virus and survival time could be noted, but the relationship was not linear on account of the excessive bias due to the mice which survived indefinitely. However, this bias can be removed by converting to reciprocals, since indefinite survival will give a reciprocal of zero, and this will not bias the mean reciprocal survival time of the group. The mean reciprocal survival times were accordingly calculated for each group of mice and plotted against the dose of compound on a logarithmic scale. A linear relationship is obtained. The horizontal line indicates the mean reciprocal survival time in the controls; the fact that the slope of the treatment line is obviously significantly different from zero proves that an antiviral effect is being obtained, and the ordinate of zero marks the dose at which all the animals survive.

Dose-response curves obtained in this manner afford a means of developing a conventional pharmacological 4-point assay of antiviral activity against a standard. All that is necessary is to infect four groups of mice with the same dose of virus, dose two groups with the compound under test given in two dose levels in a fixed ratio, and dose the other two with the standard compound given at dose levels in the same ratio.

Once in possession of a reliable assay method we were able to study the effect of modifying the molecule of isatin 3-thiosemicarbazone in various ways (Bauer and Sadler, 1960). The molecule consists of two parts, a side-chain, and a nucleus consisting of a fused benzene and pyrollidine ring. Any alteration which is made to the side-chain results in the destruction of antiviral activity. Particularly striking is the fact that activity disappears if the sulphur atom is replaced by oxygen or an imino-group. Substitution in the aromatic ring decreases activity, particularly in the 5-position and somewhat less in the 6-position.

The modifications mentioned so far have all led to a reduction or loss of activity, but when substituents are introduced in the 1-position increases in activity can be obtained. The introduction of an alkyl substituent leads to a rise in activity as far as ethyl and the activity then falls off again as the chain is lengthened. The most active compound is 1-ethylisatin 3-thiosemicarbazone, with an activity three times that of the parent compound, and the 1-methyl and 1-hydroxyethyl compounds are twice as active.

At this stage the next thing to do is to pick out a suitable derivative for clinical trial. This is not necessarily the most active compound in the series. Other factors come in, such as ease of preparation on a large scale, and lack of toxicity. However, we now come to the moment of truth for the antiviral chemotherapist. A compound has been found which has an impressive degree of antiviral activity. Have we any justification for hoping that it will have any practical application? Is there any hope that the virus disease of man will be treatable or preventable? In fact, is antiviral chemotherapy worth while?

In order to answer this question I propose to go into the question of the possibilities of the treatment and prophylaxis of virus diseases with chemical agents, at the same time summing up the position of the immunological approach, which has up to quite recently been the sole way of dealing with virus infections.

Some virus diseases can be successfully treated with drugs. These are herpetic and vaccinial keratitis, which respond to idoxuridine, and eczema vaccinatum and vaccinia gangrenosa, which respond to methisazone. We can therefore say that the principle that some virus diseases are treatable has been established. Of these diseases, we may expect some to be preventable by giving drugs, in other words, by applying chemoprophylaxis. This has been shown to be the case in the prevention of smallpox by methisazone (Bauer, St. Vincent, Kempe and Downie, 1963). Some virus diseases will probably not be treatable with drugs on account of particular features of the disease process. Of these, however, we may expect some to be preventable.

To sum up, antiviral chemotherapy may be expected to be successful in three categories of disease, those which are treatable, treatable and preventable, and preventable but not treatable.

Let us now turn to the achievements of the immunological approach. Generally speaking, virus diseases cannot be treated with vaccines and antisera. Antivaccinial gamma-globulin is of benefit in eczema vaccinatum and vaccinia gangrenosa, but its action is mainly to arrest the further spread of the disease by neutralizing virus particles which gain access to the blood stream. Although virus diseases may be considered to be untreatable by immunological methods, it is common knowledge that many of them are preventable. Immunology is therefore applicable to a single category of virus diseases only, those which are preventable but not treatable. Antiviral chemotherapy, with its promise of applicability in three categories of virus diseases, clearly offers much more in the way of opportunity.

Many virologists are of the opinion that the immunological approach to virus diseases is now subject to the law of diminishing returns. Many outstanding successes in preventive medicine have been achieved with virus vaccines. The development of vaccines against other diseases is beset with many problems, however. The virus may undergo constant antigenic variation, and exist in several serological types. The vaccine must therefore be polyvalent, and each component must be prepared from the strains which one hopes are going to prevail during the next epidemic season. The virus may also have poor antigenicity. The immunity produced by the vaccine may not be sufficiently long-lasting. Another factor is public apathy to vaccination. People in general are not very willing to have themselves and their children vaccinated against non-fatal and often relatively trivial diseases, which they expect not to get, or to get and put up with anyhow. There is also the administrative difficulty of fitting still more vaccinations into the already complicated schedule of immunizations which a child is expected to undergo in the first few years of life. Vaccines are prepared in living cells, which sometimes contain other viruses, and in using these vaccines we may be storing up trouble for the future.

Returning now to the control of virus diseases by chemical agents, let us now consider the factors which make a disease treatable or preventable.

Virus diseases are usually acute infections which are self-limiting, and there is often not enough time to institute treatment before spontaneous recovery sets in. There are, however, three factors which make a disease amenable to treatment.

(1) *The course may be relatively protracted.* This occurs in measles, where there is an initial febrile phase with Koplik spots, followed later by a rash which gradually progresses over the body; chickenpox, in which the rash appears in successive crops which may go on for a week or so; eye infections such as keratitis due to herpes, vaccinia, or zoster infection when the disease attracts attention right from the outset, skin diseases such as warts and molluscum contagiosum, which may last for months or years, and vaccinia gangrenosa, which smoulders away for several months before coming to its inevitably fatal conclusion.

(2) *The disease may have a prodromal phase* in which the sudden onset of high fever gives two or three days warning of what is to come. This occurs in measles, and particularly in smallpox.

(3) *The disease may have a biphasic course,* in which an initial

attack of influenzal illness is followed after an intermediate period of well-being by the development of severe symptoms in the second phase of the disease. This occurs in ECHO meningitis and biphasic meningoencephalitis. It may also occur in poliomyelitis.

All these types of disease are treatable, in the sense that there is plenty of time to administer an antiviral agent before the phase of virus multiplication has come to an end.

Let us now consider the factors which make a virus disease untreatable. The pathological process may be maximal or nearly so at the onset of symptoms; when the multiplication of the virus has reached its peak it is clearly too late to do anything about preventing it. This situation occurs in influenza, the haemorrhagic fevers, haemorrhagic smallpox, rabies, yellow fever, some types of poliomyelitis and infective hepatitis.

Another and unfortunately very widespread type of virus disease which is untreatable is the disease which is undiagnosable, or which can only be diagnosed long after the patient has recovered by the laborious techniques of isolating the virus and attempting to identify it by comparison with known viruses. This group includes the common cold. Also, there are many viruses, which occur particularly in the tropics, which cause fever of varying severity without characteristic signs which would enable a diagnosis to be made. Clearly, if we do not know what the patient is suffering from, we are not in a position to offer him any specific treatment. This problem can only be solved if we are lucky enough to find a compound which is active against all virus infections. However, until this improbable event comes to pass, I have said enough to show that many virus diseases are treatable in principle, and these and many others are preventable.

It was therefore with some hope of success in mind that I decided to develop the work on isatin 3-thiosemicarbazone to a point at which it could be applied to man. The main use of a compound active against the pox virus infections could undoubtedly lie in the prevention and treatment of smallpox. It was therefore important to find out whether isatin 3-thiosemicarbazone and its derivatives were active against infections with smallpox virus.

Smallpox virus exists in two variants, alastrim (or variola minor), and variola major. Both these viruses will multiply and cause a fatal encephalitis when injected intracerebrally in mice of less than 10 days of age.

In our work we mostly used mice up to 6 days old. In a typical chemotherapy test a number of litters of mice of about the same age and weight were pooled and redistributed to the mothers. All were then inoculated with the same dose of virus, which usually ranged up to 1,000 mouse LD50, or 10^7 pock-forming units in eggs. One or more litters were left as controls, and the mice in the others were treated with the compound under test in doses calculated from the body weight and injected subcutaneously in a volume of 0·01 to 0·03 cc with a micrometer syringe. The first dose was given shortly after infection and dosing was continued twice daily for five days.

TABLE 2. Effect of 1-ethylisatin 3-thiosemicarbazone in infant mice infected intracerebrally with alastrim virus

Group	Dose mg	 mg/kg	Day of death						
Control 1	—	—	3	5½	6	8	10	18	18
Treated	0·03	10	—	—	—	—	—	—	—
	0·015	5	20	—	—	—	—	—	—
Control 2	—	—	5	7	7	7	8	9	11

The result of an experiment of this kind is shown in Table 2. All the untreated animals died after an interval ranging from 3 to 18 days. The animals treated with 0·03 mg of the compound all survived, and of those treated with 0·015 mg only one died, after a prolonged period. The activity of the compound against alastrim infection in mice is thus clearly demonstrated.

In similar experiments the same compound protected at a dose level as low as 0·25 mg per kg. These low dose-levels show that the compound has a very high degree of activity against alastrim.

Similar experiments were carried out with the 1-methyl analogue (methisazone) and the parent compound. Both of these were very active, but at somewhat higher dose levels. The compounds of the isatin 3-thiosemicarbazone series are thus highly active against alastrim infection, and we next proceeded to try them out against variola major infection (Bauer, Dumbell, Fox-Hulme and Sadler, 1962). It was found that isatin 3-thiosemicarbazone protected against variola major down to 10 mg per kg, with 100 per cent mortality in the controls. The 1-methyl derivative protected down to 2·5 mg per kg, again with 100 per cent mortality in the controls. The 1-ethyl compound was also active.

The experiments in mice therefore show that isatin 3-thiosemicarbazone and related compounds exert a prophylactic effect in mice infected with both types of smallpox virus.

We therefore decided that it would be worth while to carry out a trial in man to see whether any protective effect could be observed in persons who had been in intimate contact with smallpox patients, and who were therefore exposed to the risk of developing the disease. The actual compound selected for the trial was 1-methylisatin 3-thiosemicarbazone, otherwise known as methisazone, Marboran, or compound 33T57.

A trial of this kind must be carried out in a country where smallpox is endemic. In practice, the choice lies between India, Africa and South America. We eventually settled upon Madras, in south India, a city of over two million inhabitants in which smallpox occurs in every month of the year.

The names and addresses of the smallpox patients were obtained from the admissions register of the Infectious Diseases Hospital. We then visited the homes of the patients and recorded the names and vaccination status of the close contacts, who were persons who slept in the same room as the patient. The contacts were then either offered treatment with methisazone, or left untreated. Approximately equal numbers were taken into each group. Methisazone was given by mouth twice a day, at first for four days, and later during the trial for one day. The dose for an adult was either 1½ g or 3 g, and proportionately less for children and infants.

The contacts were kept under observation, and were visited finally after 14–16 days. Cases of smallpox which occurred among them were recorded, and formed the primary result of the trial. A record was also made of the results of any vaccinations which had been carried out after contact, and of any side-effects which could be attributed to the drug.

The contacts were then divided into sub-groups according to whether they had been vaccinated in the past or not, and whether vaccination had been carried out on them after their exposure to smallpox, and, if so, whether it had been successful or not. An analysis of this kind is essential in evaluating the results of a prophylactic trial against smallpox, since susceptibility to infection is different in the various sub-groups, which should be present in the same proportions in the treated and untreated groups at the conclusion of the trial.

The results are shown in Table 3, and can be analysed into the following groups.

(1) Persons who had not been vaccinated before exposure to infection, and were not vaccinated afterwards. No case of smallpox occurred among 23 treated contacts, whereas among 16 untreated contacts there were 9 cases with 4 deaths.

(2) Persons vaccinated for the first time in their lives after exposure, in whom the vaccination failed to take. There was 1 non-fatal case among 25 treated contacts, and 4 cases, all fatal, among 20 untreated contacts.

(3) Persons vaccinated for the first time after exposure in whom the vaccination took successfully. There was 1 non-fatal case among 54 treated contacts, and 15 cases with 3 deaths among 64 untreated contacts.

Summarizing the results so far, which were mainly in infants and young children, 2 non-fatal cases of smallpox occurred among 102 treated contacts, and 28 cases with 11 deaths among 100 untreated contacts.

All these contacts lacked immunity, since primary vaccination after exposure, even if successful, does not produce immunity in time to confer protection against smallpox. The marked reduction in the incidence of cases and deaths in the treated group must therefore be ascribed to the effect of the drug.

The remaining contacts in the study had all been vaccinated one or more times in the past, before their exposure to smallpox, and therefore had a varying degree of residual immunity. They can be divided into the same three groups.

(1) Not revaccinated after exposure. No case of smallpox occurred among 95 treated contacts, and 6 cases, with no deaths, among 85 untreated contacts.

TABLE 3. Vaccination status of contacts and occurrence of contact cases

Vaccination status	Treated			Controls		
	No. of contacts	Cases of Smallpox	Deaths	No. of contacts	Cases of Smallpox	Deaths
PV at time of exposure						
Not done	23 (2·09)*	0 (0)**	0 (0)	16 (1·42)**	9 (56·2)*	4 (25·0)**
Failed	25 (2·27)	1 (4·0)	0 (0)	20 (1·78)	4 (20·0)	4 (20·0)
Successful	54 (4·90)	1 (1·8)	0 (0)	64 (5·68)	15 (23·4)	3 (4·7)
Total	102 (9·26)	2 (2·0)	0 (0)	100 (8·88)	28 (28·0)	11 (11·0)
Successful PV in infancy; RV at time of exposure						
Not done	95 (8·63)	0 (0)	0 (0)	85 (7·55)	6 (7·1)	0 (0)
Failed	546 (49·58)	1 (0·2)	0 (0)	523 (46·44)	21 (4·0)	0 (0)
Successful	200 (18·16)	0 (0)	0 (0)	257 (22·83)	15 (5·8)	1 (0·4)
Total	841 (76·38)	1 (0·12)	0 (0)	856 (76·82)	42 (4·91)	1 (0·12)
Date incomplete†	158 (14·34)	0 (0)	0 (0)	161 (14·30)	8 (5·0)	0 (0)
Overall Total	1101	3 (0·27)*	0 (0)*	1126	78 (6·92)*	12 (1·06)*

PV = Primary vaccination; RV = revaccination.
* = Percentage of overall total of contacts in the same group.
** = Percentage of number of contacts of same vaccination status.
† = The investigation was still continuing at the time of writing.

(2) Revaccinated unsuccessfully after exposure. One non-fatal case of smallpox occurred among 546 treated contacts, and 21 cases with no deaths among 523 untreated contacts.

(3) Revaccinated successfully after exposure. No cases among 200 treated contacts, and 15 cases with 1 death among 257 untreated contacts.

Summarizing these three groups of persons who had been vaccinated previously, there was 1 non-fatal case among 841 treated contacts, and 42 cases with 1 death among 856 untreated contacts.

The incidence of smallpox is thus lower in persons who have been vaccinated (4·91 per cent, compared with 28 per cent), but is far from being zero. The protective effect of the drug is apparent in these people also.

Summarizing the total results of the trial, including persons of all types of vaccination status, there were 3 non-fatal cases among 1,101 treated contacts, and 78 cases with 12 deaths among 1,126 untreated contacts. Treatment with methisazone may therefore be assumed to have prevented some 70–80 cases of smallpox with their associated deaths.

The results also show the effect of vaccination carried out after exposure. Vaccination, whether successful or not, was carried out on 825 persons in the treated group, among whom there were 3 non-fatal cases of smallpox; among the control group 864 persons were vaccinated, and of these 55 developed smallpox and 8 died. Vaccination carried out after exposure will therefore not prevent the occurrence of contact cases of smallpox, and judged as a public health procedure for use in the field, prophylactic treatment with methisazone is a more effective procedure.

It has thus been shown that a virus disease of man is preventable by drug treatment, and antiviral chemoprophylaxis is therefore an established fact.

Turning now from prevention to the treatment of established disease, this is also an accomplished fact. Herpetic keratitis in man can be cured with idoxuridine. Idoxuridine is the approved name of 5-iodo-2′-deoxyuridine, an antimetabolite which prevents the incorporation of thymidine into deoxyribonucleic acid. It will inhibit the multiplication of herpes and vaccinia viruses in tissue culture, and is now widely used in the treatment of herpetic ulcers of the cornea. Acknowledging that idoxuridine is an effective agent in the treatment of an established virus disease, I do not propose to discuss it further, as I cannot write from personal experience, and as an example of the chemotherapy of established virus infection I will discuss the use of methisazone in the treatment of complications of smallpox vaccination.

Post-vaccinial encephalitis does not respond to chemotherapy; this is not surprising, since the condition is probably allergic in origin, and not due to multiplication of the virus in the brain.

Some apparent success has been achieved in the treatment of eczema vaccinatum. Thus, methisazone appeared to have been effective in 12 of 22 cases. The course of eczema vaccinatum is so variable, however, that one is forced to rely on clinical impressions in assessing the results of treatment, and it would be difficult to accumulate enough cases for a double-blind trial as the disease is relatively uncommon.

One is on firmer ground in the treatment of vaccinia gangrenosa, since this complication of vaccination has been invariably fatal. Recovery has been obtained in 5 of 10 patients with vaccinia gangrenosa treated with methisazone.

I think that I have said enough to show that antiviral chemotherapy and chemoprophylaxis are established facts. Smallpox can be prevented with methisazone, herpetic keratitis can be treated with idoxuridine, and vaccinia can be treated with methisazone. These drugs are in use at the present time.

While the further outlook gives reason for optimism, there are difficulties facing the future development of antiviral drugs. The main problem to be solved is the narrow spectrum of activity. Methisazone is only active against certain of the pox viruses. Idoxuridine includes herpes in its spectrum of activity. It is rather as if penicillin were only active against streptococci, and drugs affecting staphylococci, pneumococci and other common organisms had not yet been discovered.

The prospects for discovering a drug active against most, or all, viruses are not very encouraging. An antiviral drug should act against the virus, not against the host cell; and there is yet no reason to suppose that viruses have sufficient in common at the chemical level to make them all susceptible to the same chemical agent.

However, I am quite certain that the antiviral drugs which will be used in ten years' time will no longer be methisazone or idoxuridine. They are merely the first means of entry into a hitherto closed field of therapeutics, and we can confidently expect a series of developments which will parallel the rise of the sulphonamides after the initiation of the era of antibacterial thermotherapy.

REFERENCES

Bauer, D. J. (1955) *Brit. J. exp. Path.*, **36**, 105.
Bauer, D. J., Dumbell, K. R., Fox-Hulme, P. and Sadler, P. W. (1962) *Bull. World Health Org.*, **26**, 727.
Bauer, D. J. and Sadler, P. W. (1960) *Brit. J. Pharmacol.*, **15**, 101.
Bauer, D. J., St. Vincent, L., Kempe, C. H. and Downie, A. W. (1963) *Lancet*, **ii**, 494.
Domagk, G., Behnisch, R., Mietsch, F. and Schmidt, H. (1946) *Naturwissenschaften*, **33**, 315.
Hamre, D., Bernstein, J. and Donovick, R. (1950) *Proc. Soc. exp. Biol. Med.* **73**, 275.
Thompson, R. L., Minton, S. A., Officer, J. E. and Hitchings, G. H. (1953) *J. Immunol.*, **70**, 229.

XII

Dilute and Concentrated Urine

J. N. MILLS

OUR understanding of the means whereby the mammalian kidney elaborates concentrated urine was obscured for many years by certain preconceptions which, together with a lack of imagination, acted as blinkers to our vision. Fifteen years ago it was well known that the great bulk of the fluid filtered at the glomerulus was reabsorbed further down the nephron leaving, in man, at one extreme some 20 ml of very hypotonic urine or at the other a fraction of a ml of highly concentrated urine with osmotic pressure up to four times that of the plasma. Some mammals can achieve higher concentrations than this: rats up to nine times, and some desert rodents up to twenty times the plasma osmotic pressure.

The elaboration of hypotonic urine was ascribed with fair plausibility to the active reabsorption of solute, such as sodium, through a membrane more or less impermeable to water. The elaboration of hypertonic urine was similarly ascribed to an active reabsorption of water. It was also known that the posterior pituitary antidiuretic hormone (ADH) was necessary for the elaboration of hypertonic urine; and Verney's painstaking experiments had shown that small injections of hypertonic fluid into the blood supplying the supraoptic nucleus could induce ADH secretion, whilst similar injections elsewhere were without effect. The posterior pituitary appeared to be tonically active and could be depressed by hypotonicity in the supraoptic nucleus. It seemed that ADH must exert two distinct actions; the conversion of hypotonic to isotonic urine necessitated only the permission of water movement down an osmotic gradient, as could be achieved by some such process as the opening of water-permeable pores; the further conversion of isotonic to hypertonic fluid needed, it seemed, a stimulation of active water transport.

Early this century it was known that the thin segment of Henle's loop was found only in mammals and birds, the only vertebrates capable of elaborating hypertonic urine, so the loop was believed to be responsible for this renal activity. 'Responsibility' came to be accepted as implying the active abstraction of water from the lumen. This conception was shattered when Walker and Oliver (1941) published the first accounts of mammalian tubular puncture and showed that samples from the distal convoluted tubule were hypotonic even though the urine was hypertonic. The collecting ducts have a structure simpler than that of the convoluted tubules so were commonly held to be passive conduits; it seemed that if the limited number of analyses by Walker and Oliver were reliable the active water reabsorption must be in the further part of the distal convolution inaccessible to micropuncture. The suggestion that Henle's loop was concerned with elaboration of hypertonic urine was generally abandoned and a variety of somewhat naïve speculations about the function of the loop were substituted.

The interpretation of hypotonicity as due to active reabsorption of solute from a waterproof segment still remains acceptable, and this process extends into the collecting duct (Hilger, Klümper and Ullrich, 1958). So far as the mechanisms for elaboration of hypertonic urine are concerned, the facts so far recorded have hardly needed modification, but our present-day conclusions are widely different. It is worth pausing for a moment to consider the fallacies in thought which for so long delayed advance in our understanding. These were primarily two. First, the structural arrangement of the nephron was largely ignored; it was treated functionally as if it might be unravelled and stretched out, being coiled and kinked within the kidney merely for compactness, and no relevance was seen in the proximity of Henle's loop, the collecting ducts and the vasa recta. The second fallacy was the uncritical acceptance of Claude Bernard's concept of the constancy of the milieu interieur as being uniform all over the body.

The idea of a counter-current multiplier, developed by a physical chemist Kuhn (Wirz, Hargitay and Kuhn, 1951), was wrong in much physiological detail. I propose therefore to develop the concept not in historical order but as it might have evolved in the mind of a physiologist more perceptive than were any of us at the time.

If hypotonic fluid leaves the distal convolution and strongly hypertonic fluid reaches the renal pelvis the implied abstraction of water could be achieved by known forces if the medullary tissues were hypertonic. Only the assiduous bibliophil would have known of Hirokawa's (1908) measurements which demonstrated that this is, in fact, the case, but any careful attempt to measure osmotic pressure or concentrations of solutes in the medulla would have demonstrated this, as shown by Wirz *et al.* (1951). How could such an aberrant milieu persist while elsewhere the circulation of the blood preserves a constant milieu? Reflection would have shown that the hairpin arrangement of the vasa recta, the only vessels supplying the medulla, would minimize the dissipation of any substance accumulated in the medulla. In so far as the ascending vessels carried away any diffusible solute, this would tend to diffuse back into the lower concentration in the descending vessels and hence to return to the medulla. A similar counter-current exchange of heat in the limb vessels had long been familiar to physiologists.

The previous suggestion that the loop of Henle was responsible for urinary hypertonicity would immediately fit into place if one supposed that the loop was responsible for generating this interstitial hypertonicity, which in due course resulted in water abstraction from the collecting duct. If the loop were capable of active abstraction of solute from the lumen, such an interstitial hypertonicity could readily be produced, and the known hypotonicity of the fluid reaching the distal tubule would likewise be explained. The only necessary physiological mechanism would be a sodium pump such as was known to exist in many tissues.

The argument has thus far been developed without considering the operation of a counter-current multiplier; this was the essential contribution of the original paper of Wirz *et al.* (1951) but it is not an easy concept at first acquaintance, and I have therefore developed the argument for the means of production of urinary hypertonicity up to this point without mentioning it.

Fundamentally, the concept is one whereby small transverse uphill transport can lead to large longitudinal gradients, so that a high degree of hypertonicity can be developed at the tip of the medulla by the operation of a pump which can only work against a limited electrochemical gradient. Suppose that the ascending limb of Henle's loop contains a sodium pump of this sort, and that the descending limb is permeable to water or solute or both. Fluid descending into the medulla will thus encounter a region of increased osmotic pressure and thus itself become passively concentrated. As this concentrated fluid doubles back round the loop, and solute extraction continues, the interstitium near the tip of the loop will be further concentrated. Further proximal fluid approaching the tip will thus attain a higher osmotic pressure before doubling back round the bend; further active abstraction of solute will then further concentrate the fluid in the interstitium. I have elaborated this elsewhere in a diagrammatic presentation (Mills, 1963).

What evidence is there for this picture? The osmotic pressure is high, and equally high, in samples of fluid collected from the loop of Henle, the vasa recta and the collecting duct at the same level (Gottschalk, 1964). The ascending limb is relatively impermeable to water while the descending limb is freely permeable (Gottschalk, 1963), and preliminary results suggest that the osmotic pressure of fluid in the ascending limb may be significantly below that of other structures, or at least of fluid in the collecting duct, at the same level. The evidence is not complete, but the general picture of hypertonicity achieved by extrusion of sodium from a waterproof ascending limb is fairly clear.

How is it that this medullary hypertonicity is not swamped by the water passively abstracted from the collecting duct in the final concentration of the urine? The explanation for this depends upon the operation of the distal convolution, surrounded by a normal network of capillaries which presumably maintain a milieu resembling that of the rest of the body; here the volume of fluid is reduced from some 20 per cent to 5 per cent of that filtered at the glomerulus (Lassiter, Gottschalk and Mylle, 1961). It may be that the volume is further reduced in the inaccessible part of the tubule before the medulla is reached.

Much earlier work, which has received abundant recent confirmation (Levinsky and Berliner, 1959), indicates that higher urinary osmolalities can be attained if urea contributes substantially to the total solutes. The likely explanation is that the walls of the collecting duct are moderately permeable to urea which, as it is concentrated by water abstraction, diffuses back from the lumen into the interstitium down a concentration gradient. The high urea concentration in the lumen thus does not contribute to the effective osmotic pressure of its contents. Sodium and urea are, in fact, the solutes mainly responsible for the high osmotic pressure in the medulla.

How far is this mechanism modified in a water diuresis? This is a more difficult subject for study, since most anaesthetics cause liberation of ADH. Many investigations have therefore been performed upon animals with lesions in the hypothalamus producing experimental diabetes insipidus; the difficulty here is the uncertainty whether any other function has been altered by these lesions. In either condition, water loading or experimental diabetes insipidus, the medulla remains hypertonic but its osmotic pressure is much lower than in hydropenia. Since the urine in the collecting duct is very hypotonic there is an obvious failure of osmotic equilibration across its walls; it has therefore been suggested that ADH primarily increases its water permeability, as has been also suggested in the toad bladder (Orloff and Handler, 1964). This is a simple but inadequate explanation for the action of ADH, since hypertonic urine can still be formed in its absence, if, for instance, glomerular filtration is restricted (Berliner and Davidson, 1957). If we examine the composition of fluid collected from the distal convoluted tubule, we find that in the presence of ADH (Gottschalk and Mylle, 1959) the hypotonic fluid entering this convolution attains isotonicity; the walls appear to be water-permeable. By contrast, in water diuresis the fluid in this convolution remains hypotonic (Wirz, 1956) so that the walls are obviously relatively impermeable to water. It appears that this is an important site of action of ADH which, by causing the passive loss of water from the distal tubule down an osmotic gradient, grossly reduces the volume of fluid delivered to the collecting duct. In the absence of ADH the volume flowing down the collecting duct may be so large that osmotic equilibration is no longer possible, and it may well be that water reabsorption from this structure is greater in water diuresis than in antidiuresis. It thus no longer seems necessary to suppose that ADH increases the water permeability of the collecting duct, though it may well do so. Further evidence that ADH acts primarily upon the distal tubule comes from radioautographs after the administration of ADH labelled with isotopic iodine on the tyrosine, which show it to be fixed primarily in the distal convolution (Darmady, Durant, Matthews and Stranack, 1960). Direct measurements of water permeability of the distal convolution under a known osmotic gradient indicate that it is increased some 2·3 times by ADH (Ullrich and Rumrich, 1963).

Any increase in medullary circulation would tend to wash away the hyperosmolality, and the blood flow is said to increase in water diuresis (Thurau, Deetjen and Kramer, 1960) which could provide a physiological role for the known vasoconstrictor effect of ADH, formerly known as vasopressin; however, it is said that physiological doses of ADH do not detectedly restrain medullary blood flow (Thurau, 1964).

If small samples of medullary tissue are analysed for their separate constituents, in conditions of water diuresis and hydropenia, it is found that in water diuresis there is somewhat more water, considerably less urea, and substantially the same amount of sodium (Saikia, 1965*a*; Fig. 1). Since the water content, which most workers have not measured, varies, one must adopt some basis for the computation of amounts of water or solute in a given amount of tissue. Dry weight is an unsuitable basis since the amounts of urea vary sufficiently widely to contribute significantly to the dry weight. The amounts shown in Fig. 1 are therefore calculated on the basis of urea-free dry weight, which is closer to a measure of the structural solids of the cell. The altered amounts of these components lead to the change in composition of different medullary segments shown in Fig. 2.

Administration of ADH in doses believed to be physiological, which reduce urine flow to a low level, completely fail to reproduce the medullary composition found in hydropenia (Saikia, 1965*b*). It is also well known that such injection fails to

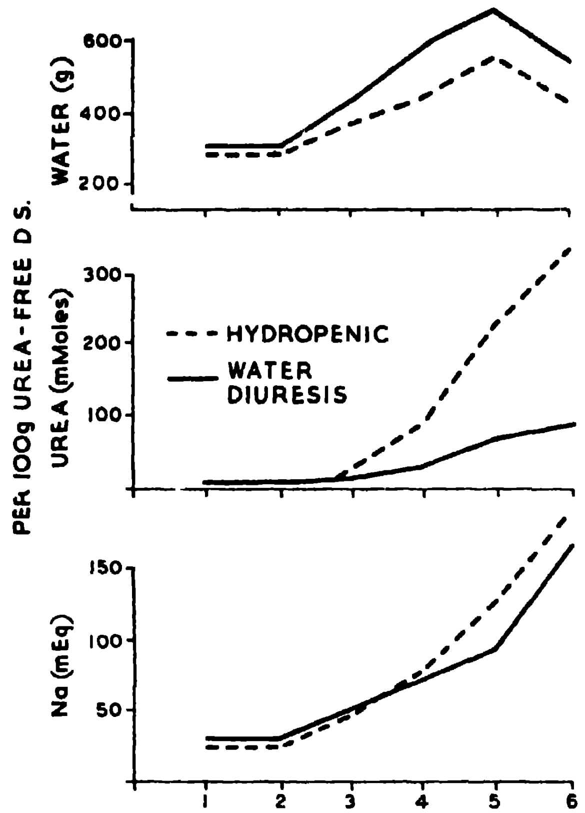


Fig. 1. Water, urea, and sodium content of slices from kidney of hydropenic and water diuretic rats, numbered as in Fig. 2.

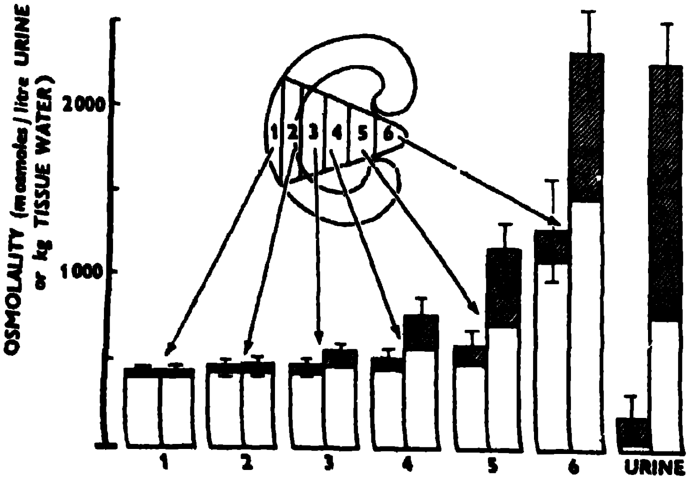


Fig. 2. Mean osmolality of slices from rat kidney (as shown in the inset) and urine, with S.D. Right-hand column of each pair represents 10 hydropenic rats, and left-hand column 10 rats in water diuresis. In each rectangle, clear part represents m Osmoles contributed by electrolytes

$$(Na + K + NH_4 + \text{equivalent anions}),$$

and shaded part m Osmoles contributed by urea. (Reproduced by permission from Saikia, T.C. (1965) *Quart. J. exp. Physiol.*, 50, 153.)

achieve the maximal urinary osmolality which can be attained after a period of hydropenia. It appears, then, either that hydropenia affects renal function by some other means additional to that due to ADH, or that prolonged exposure to ADH has effects beyond those which can be exerted acutely. Some evidence for the former possibility comes from the ability of dogs with diabetes insipidus to produce moderately hypertonic urine if deprived of water, although the maximally attainable osmolality is, of course, far below that in intact animals.

It seems then that the acute effect of ADH is on permeability, increasing water loss from the distal convolution and probably also urea loss from the collecting duct. This latter action has been demonstrated by direct micro-perfusion (Gardner and Maffly, 1964); and in man (Thomas, 1964) and in the rat (Saikia, 1965*b*) ADH injection causes a transient drop in urea excretion which is most easily explained by loss from the collecting duct into the renal medulla. Experiments on the toad bladder (Orloff and Handler, 1964) suggest that all the actions of ADH are passive, permitting more rapid downhill penetration of water and urea, and accelerating sodium transport only at a barrier where movement is down an electrochemical gradient. There have been a number of attempts to explain this increased permeability. The claim of Ginetzinzky that the hormone causes an increased liberation of hyaluronidase has now little experimental support. It has, however, been clearly shown that ADH stimulates the enzymic conversion of adenosine triphosphate to adenosine 3′-5′-monophosphate (cyclic AMP), that theophylline inhibits the enzymic destruction of cyclic AMP, and that theophylline or cyclic AMP can reproduce all the permeability changes in toad bladder ascribed to ADH. It is thus postulated (Orloff and Handler, 1964) that cyclic AMP is an essential intermediary in the operation of ADH, though its own mode of action upon cellular barriers is still unexplained. Among other problems still without coherent explanation is the old observation that cortisol is necessary for the development of a water diuresis, and that renal concentrating ability is impaired in potassium deficiency and in calcium excess. The only progress in explanation of these findings lies in the demonstration that in both conditions (Manitius, Levitin, Beck and Epstein, 1960*a*;

Manitius, Levitin, Beck and Epstein, 1960*b*) the medullary osmolality is much less than in normal animals.

REFERENCES

Berliner, R. W. and Davidson, D. G. (1957) *J. clin. Invest.*, **36**, 1416.

Darmady, E. M., Durant, J., Matthews, E. R. and Stranack, F. (1960) *Clin. Sci.*, **19**, 229.

Gardner, K. D. and Maffly, R. H. (1964) *J. clin. Invest.*, **43**, 1968.

Gottschalk, C. W. and Mylle, M. (1959) *Amer. J. Physiol.*, **196**, 927.

Gottshalk, C. W. (1963) *Harvey Lectures*, **58**, 99.

Gottschalk, C. W. (1964) *Amer. J. Med.*, **36**, 670.

Hilger, H. H., Klümper, J. D. and Ullrich, K. J. (1958) *Arch. ges. Physiol.*, **267**, 218.

Hirokawa, W. (1908) *Beitr. chem. Physiol. Path.*, **11**, 458.

Lassiter, W. E., Gottschalk, C. W. and Mylle, M. (1961) *Amer. J. Physiol.*, **200**, 1139.

Levinsky, N. G. and Berliner, R. W. (1959) *J. clin. Invest.*, **38**, 741.

Manitius, A., Levitin, H., Beck, D. and Epstein, F. H. (1960*a*) *J. clin. Invest.*, **39**, 684.

Manitius, A., Levitin, H., Beck, D. and Epstein, F. H. (1960*b*) *J. clin. Invest.*, **39**, 693.

Mills, J. N. (1963) In *Recent Advances in Physiology*, ed. R. Creese, 8th edn., p. 260, Churchill: London.

Orloff, J. and Handler, J. S. (1964) *Amer. J. Med.*, **36**, 686.

Saikia, T. C. (1965*a*) *Quart. J. exp. Physiol.*, **50**, 146.

Saikia, T. C. (1965*b*) *Quart. J. exp. Physiol.*, **50**. 158.

Thomas, S. (1964) *J. clin. Invest.*, **43**, 1.

Thurau, K., Deetjen, P. and Kramer, K. (1960) *Arch. ges. Physiol.*, **270**, 270.

Thurau, K. (1964) *Amer. J. Med.*, **36**, 698.

Ullrich, K. J. and Rumrich, G. (1963) *Arch. ges. Physiol.*, **44**, 278.

Walker, A. M. and Oliver, J. (1941) *Amer. J. Physiol.*, **134**, 562.

Wirz, H., Hargitay, B. and Kuhn, W. (1951) *Helv. physiol. pharmacol. Acta*, **9**, 196.

Wirz, H. (1956) *Helv. physiol. pharmacol. Acta*, **14**, 353.

XIII

Long-term Dialysis as a Substitute for Kidney Function

STANLEY SHALDON

THE treatment of chronic uraemia by repetitive dialysis can now be considered as an established entity in therapeutic medicine. Three patients who commenced treatment five years ago are still alive and well (Scribner, 1965*a*) and there are ten patients alive and well in their third and fourth years of treatment at several dialysis centres. Nevertheless, considerable controversy still exists both inside and outside medical centres concerning this therapy (Elkinton, 1964; Robin, 1964; Scribner, 1964). The major issues now raised by chronic dialysis therapy are economic however, rather than technical or medical. The problem may reflect in part the future problems that will face mankind in relation to medical treatment as medical science advances.

Chronic dialysis therapy can be divided into: (*a*) Haemodialysis; (*b*) Peritoneal Dialysis. *Chronic Haemodialysis* is undergoing a period of rapid development which has been necessitated by the urgent need to make this treatment available to all who need it by simpler more effective and less expensive methods.

TECHNICAL ASPECTS

The technical aspects of treatment can be considered under the following headings: (1) Dialyser design; (2) Dialysate supply and monitoring; (3) Cannulas.

(1) *Dialyser design*

Although many forms of artificial kidney have been used in relation to chronic haemodialysis including the twin coil Kolff disposable dialyser (Barber, Nakamoto, McCormack and Kolff, 1963; Gonzales, Pabico, Brown, Maher and Schreiner, 1963; Shaldon, Rosen and Silva, 1963) and the Skeggs-Leonard dialyser (Pendras, Cole, Tu and Scribner, 1961), results have only been satisfactory since the introduction of the two-layered polypropylene Kiil Dialyser (Cole, Pollard and Murray, 1963). The ideal dialyser design for chronic haemodialysis has to be considered in relation to four aspects. (*a*) Priming blood volume; (*b*) Intrinsic resistance to blood flow across the dialyser; (*c*) Preparation of dialyser; (*d*) Membranes.

(*a*) *Priming volume.* It is essential for long term haemodialysis that the dialyser is not primed with blood and the blood volume of the two-layered Kiil dialyser is only 400 ml, compared with 800–1,000 ml of the twin coil dialyser. Furthermore, only 20–40 ml of blood are lost in returning the blood content to the patient at the end of dialysis (Cole *et al.*, 1963). It is thus possible to dialyse patients twice weekly without donor blood priming requirements although transfusion requirements for maintenance of adequate haemoglobin levels average two pints of blood per month.

(*b*) *Intrinsic resistance.* The intrinsic resistance of the two-layered Kiil dialyser permits a passive blood flow through the unit of 200 ml per minute without the use of a blood pump. No other dialyser of the same surface area ($0{\cdot}9M^2$) achieves the same low intrinsic resistance. The Kolff twin coil unit requires a blood pump and the Skeggs-Leonard unit although permitting passive flow has a higher resistance. In chronic dialysis it is undesirable to use a blood pump routinely because of increased haemolysis. However, adequate monitoring control does permit the occasional use of a blood pump to boost passive flow without risk of bursting the dialyser from a build up of pressure within the blood compartment.

(*c*) *Preparation of dialyser.* The preparation of the dialyser unit unless it is sterile and disposable, limits its use. Although the two-layered polypropylene Kiil dialyser has to be rebuilt for each use, the time consumed by skilled personnel is not excessive. The boards of the dialyser are scrubbed in a detergent solution and then the clean boards are assembled with cuprophan membranes, pressure tested and filled with 1 per cent formalin in 30 minutes. Air venting is not required if after air pressure testing, carbon dioxide is flushed through the blood compartment, before the unit is sterilized in formalin.

(*d*) *Membranes.* The potential of a flat dialyser lies in its use of flat sheet membranes as opposed to coiled tubes which require extrusion in tube form. The major advantage of the two-layered Kiil dialyser over the twin coil unit is because thin cellophane PT150 (cuprophan) can be used. This latter membrane permits faster clearances of urea, uric acid and creatinine than the PT300 cellophane used in twin coil units (Lyman, 1963; Smith, 1965).

(2) *Dialysate supply and monitoring*

Initially dialysate supply was refrigerated because of the bacterial growth rate in dialysate prepared dilute in static tanks at 40° C. when the dialysate was returned to the static tank after passage through the dialyser (Sherris, Cole and Scribner, 1961). However, single pass of the dialysate to waste after traversing the dialyser has reduced the contamination rate within the static tank and eliminated the need for refrigeration. This has also eliminated the need to rewarm the blood prior to its return to the patient thus shortening the blood line from the dialyser to the patient. In addition, urea clearance across a cellophane membrane is increased by 30 per cent if the dialysate temperature is raised 30° C. (i.e. from 7° C.–37° C.) (Shaldon, Silva and Rosen, 1964). The improved blood flow resulting from the reduction in the blood line and the higher urea clearance have increased considerably the efficiency of the two-layered polypropylene Kiil dialyser. Furthermore, with 37° C. single pass dialysis the optimum dialysate requirement has dropped to 500 ml per minute. On a small scale it is thus possible to prepare enough dialysate supply for one 14 hour haemodialysis in a 500 litre static tank. This system can easily be cleansed but bulk preparation and storage of dialysate fluid on a large scale has limited the expansion of the treatment. To overcome this problem, and reduce the time and cost of manufacturing dialysis fluid, a proportionality pumping system using liquid dialysate concentrated 35 times and mains water supply has been developed (Grimsrud, Cole, Lehman, Babb and Scribner, 1964). The system is fully automated and includes a self-sterilizing cycle and conductivity metering to ensure the correct sodium concentration of the dialysate. Analytical errors in the concentrate preparation are compensated for by feed back control from the conductivity meter to the proportioning pumps. The system can supply multiple hospital dialyser units or single units for home use. The dialysate composition has been standardized in relation to sodium acetate which has replaced sodium bicarbonate in dialysis fluid as sodium bicarbonate cannot be concentrated successfully in liquid form. Initially, monitoring of dialysis was performed by trained nursing staff who regularly took the patient's blood pressure and observed the blood flow. Incorporation of sensitive Bourdon Pressure Gauges with adjustable electrical alarm contacts have overcome all difficulties in relation to unobserved monitoring of overnight dialysis. Furthermore, active patient participation in dialysis has greatly reduced the need for trained nursing observation. The air pressure monitor recording pressure changes in an enclosed plastic chamber on the return line of blood from the dialyser to the patient can detect alteration in the patient's blood pressure, blood leaks in the dialyser or obstruction to flow from fibrin deposition. This monitor can be incorporated into an alarm system to cut out a blood pump if it is in use, or the effluent pump sucking dialysate through the dialyser. This latter procedure is also monitored by a second Bourdon gauge with a vacuum pressure range 0–250 mm Hg, so that water removal from the patient can be controlled by adjustable negative pressure applied to the dialysate compartment of the dialyser. The second gauge is also wired to an alarm cut out system to ensure against build up of negative pressure from air locks in the feed line from dialysate supply source to the dialyser. Weight removal from the patient during dialysis is monitored by a bed weigher recording losses of as little as 25 g. Using these monitoring devices, the aims of self-service dialysis in hospital and overnight unobserved haemodialysis in the home have been realized.

(3) *Cannulas*

The major advance in chronic haemodialysis came with the introduction of the silastic (silicone rubber) teflon arterio-venous shunt (Quinton, Dillard, Cole and Scribner, 1962). This prosthesis is exteriorized (Plate IX, Fig. 1)[1] and the artificial arterio-venous shunt is broken every time dialysis occurs. To facilitate self-dialysis, the leg cannula sites are preferred and a modified exterior portion has been introduced by the elimination of the teflon by-pass unit with a single make and break junction (Sevitt, Comty, Rottka and Shaldon, 1964). Complications of cannula insertion and long-term function include infection, clotting, accidental opening of the make and break joint and traumatic damage. To date none of these problems have been insuperable and with adequate care and attention individual cannulas have functioned satisfactorily for over 12 months and the longest functioning cannula has been in place for over 3½ years. Exhaustion of cannula sites would not seem to be a critical factor for periods of 5–10 years, but ultimately may become the limiting factor for survival beyond ten years.

Home haemodialysis

The application of dialysate supply advances and monitoring control have made overnight haemodialysis in the home a reality. To date, two patients with trained relatives have been dialysing satisfactorily in the home overnight whilst asleep for six and two months respectively (Baillod, Comty, Ilahi, Konotey-Ahulu and Shaldon, 1965). Day haemodialysis in the home has also been practised in two centres (Curtis, Cole, Fellows, Tyler and Scribner, 1964; Merrill, Schupak, Cameron and Hampers, 1964).

Peritoneal dialysis

The intermittent use of peritoneal dialysis has proven of extreme value in selected patients. One patient has survived 24 months (Boen, 1965) and many patients have survived over 12 months. The abandonment of an indwelling prosthesis in the peritoneal cavity and the replacement of the catheter on each dialysis with its removal at the end of dialysis have eliminated infection risks and reduced protein loss in the dialysate to less than 20 g per dialysis. The use of 3 litres per hour of dialysate exchange with a non-irritant chemical formulation containing sodium chloride 95 mEq per litre, sodium acetate 35 mEq per litre, calcium chloride 3 mEq per litre, and magnesium chloride 1 mEq per litre, dextrose 1·7 g per 100 ml, have resulted in relatively pain free dialysis without protein depletion. In fact, patients have gained flesh in our experience. Automatic cycling of the dialysis fluid has turned peritoneal dialysis into an entirely automated procedure (Boen, Mion, Curtis and Shilipetar, 1964). The potential of peritoneal dialysis in long-term treatment has yet to equal the results of haemodialysis, but the results are sufficiently good to warrant further exploration of the technique. In addition, it is an excellent reserve holding therapy when cannula problems or haemorrhagic complications of uraemia contraindicate haemodialysis treatment temporarily.

---

[1] This plate will be found facing p. 210.

Medical Aspects

Twenty patients (13 males and 7 females, ages 18–57 years) are currently under treatment at the Royal Free Hospital Renal Unit (Table 1). Fifteen patients are receiving twice weekly 14 hour overnight haemodialysis in hospital and two are receiving the same treatment at home. Three patients are receiving once weekly 16–20 hour peritoneal dialysis therapy. The longest survivor is now in his third year. All patients were suffering from end stage renal failure due to chronic glomerulonephritis (11 patients), chronic pyelonephritis (7 patients) and polycystic kidneys (2 patients).

Patient selection

Criteria for patient selection are becoming less stringent as a result of technical advances. Nevertheless, patient co-operation in diet, cannula care, and ability to adjust to regular therapy are essential for successful results. Emotionally mature and stable adults still make the best patients. Patients suffering from multi-system diseases associated with renal failure (e.g. disseminated lupus erythematosus, amyloidosis, polyarteritis

TABLE 1. Chronic dialysis

Patient		Age	Disease	Rehabilitation	Duration of treatment (months)	24 hour urine volume
			Haemodialysis			
E.R.	M	24	G.N., M.H.	100%	27	0
R.J.	M	34	G.N.	100%	21	0
O.H.	F	34	Pyelo., M.H.	100%	19	0
P.J.	M	42	Polycystic	100%	16	300
R.B.	M	24	G.N., M.H.	100%	16	0
K.H.	M	33	G.N., M.H.	100%	16	100
J.B.	M	38	G.N.	100%	12	50
B.C.	M	31	Pyelo. 1 Gout	100%	11	100
J.W.	M	36	G.N., M.H.	80%	7	100
D.M.	M	25	G.N., M.H.	100%	7	150
M.S.	M	20	Pyelo.	80%	7	0
W.M.	M	29	Pyelo.	100%	5	50
R.H.	F	26	G.N., M.H.	50%	5	50
M.F.	M	26	G.N., M.H.	50%	3	0
H.M.	M	38	Pyelo.	75%	3	300
P.W.	F	36	G.N., M.H.	50%	3	0
A.W.	F	32	G.N., M.H.	50%	2	0
			Peritoneal dialysis			
N.C.	F	57	Polycystic	80%	12	100
C.P.	F	18	Pyelo.	80%	5	50
M.K.	F	40	Pyelo.	80%	4	100

M.H. = Malignant hypertension.
G.N. = Glomerulonephritis.
Pyelo. = Pyelonephritis.

nodosa) are not considered ideal candidates for therapy although one patient with disseminated lupus erythematosus and renal failure has been under treatment successfully in Seattle for one year (Curtis, Cole, Fellows, Tyler and Scribner, 1965). Our patients are all suffering from primary renal failure but in addition, one patient (B.C.) has primary gout. Malignant hypertension has not proven to be a contraindication to successful therapy, but patients with evidence of advanced vascular disease (e.g. previous myocardial infarction, or cerebral thrombosis) associated with long-standing hypertension should be excluded at the moment. For this reason, an arbitrary age limit of 55 years has been imposed on all patients with severe hypertension. However, patient N.C. aged 57 has survived one year on peritoneal dialysis but she had only a moderate degree of hypertension at the start of treatment. In addition, adolescents and children are unable to cope with the demands of the system and the lower limits of age inclusion has been defined as 18 years.

Rehabilitation (Table 1)

Rehabilitation has been estimated at 100 per cent when the patient is able to work a full week. This usually takes 3–6 months to achieve from starting treatment. Arduous physical activity is limited by anaemia and the need to restrict the use of the cannulated limb, nevertheless, patients have returned to various occupations including factory work, chartered accountancy, business administration, play writing, shopkeeping and housekeeping. Two male patients have married since starting treatment and one has procreated. Females remain infertile but males retain potency. Libido returns in all patients. The problems of rehabilitation should not be minimized, they involve constant driving of the patients to greater self-effort and have much in common with the rehabilitation of any disabled person. Psychological stresses are greatest in the first six months of treatment but become less overt with adaptation to a new way of life and acceptance of their existence and continued survival by family, friends and employers. Self-service dialysis in hospital and home dialysis have done much to eliminate the psychological stress associated with treatment.

Mortality

Two patients have died in the past 24 months, but there have been no deaths in the last 16 months. Table 2 details survival

TABLE 2. Mortality

Patient	Age	Disease	Duration of treatment (months)	Cause of death
P.P.	33	Pyelo. MH	10	pulmonary embolus
E.H.	39	GN MH	7	mesenteric thrombosis

No deaths in past 16 months since using arteriovenous cannulae and 2 layered Kiil dialyser.

times and causes of death in those two patients. Patient P.P. died from a pulmonary embolus associated with indwelling femoral venous cannulas (Shaldon *et al.*, 1963) before the adoption of the silastic teflon arterio-venous shunt (Quinton *et al.*, 1962). Patient E.H. had hypertensive heart disease for seven years prior to the onset of treatment and at post-mortem had advanced atherosclerosis.

Uraemic control (Table 3)

Biochemical control by dialysis is detailed in Table 3. Patients are maintained completely free from uraemic symptoms and gain flesh if 14 hour dialysis is given twice weekly using a cuprophan membrane. Nevertheless, an almost comparable state of well-being has been achieved with once weekly peritoneal

TABLE 3. Average plasma urea, creatinine, uric acid and potassium levels before and after dialysis

		PLASMA							
	Body weight	Urea mg%		Creatinine mg%		Uric acid mg%		K+ mEq/l	
		Before	After	Before	After	Before	After	Before	After
Haemodialysis—14 hours twice weekly									
R.J.	76	120	40	11·5	4·1	8·0	2·0	6·0	4·0
K.H.	66	140	45	13·0	5·0	8·1	2·0	6·1	3·9
E.R.	55	110	29	11·0	3·8	7·5	1·8	6·5	4·0
P.J.	62	150	50	9·1	2·0	6·5	2·0	5·2	3·8
R.B.	62	85	20	10·0	4·1	7·8	3·1	5·7	3·9
O.H.	45	100	24	7·2	2·1	6·4	1·6	6·0	4·0
J.B.	63	110	40	11·5	4·7	9·1	2·3	6·1	4·1
B.C.	66	130	45	12·0	5·0	10·1	4·7	5·0	3·5
J.W.	56	140	55	13·0	4·2	8·0	5·0	6·5	3·9
D.M.	49	100	30	11·0	5·0	7·5	4·0	5·8	3·9
M.S.	55	95	25	10·0	3·1	8·0	2·0	5·5	4·0
W.M.	61	120	40	12·0	4·9	8·5	3·5	5·8	4·3
R.H.	44	100	30	11·0	3·0	7·0	2·0	5·0	3·5
M.F.	55	110	30	10·0	2·9	6·5	1·9	5·2	3·6
H.M.	71	120	45	11·2	3·8	6·8	3·0	4·9	4·1
P.W.	49	100	20	10·1	4·0	7·1	2·5	5·1	3·9
A.W.	56	140	40	9·7	3·1	7·2	2·3	6·0	3·7
Peritoneal dialysis—16–20 hours once weekly									
N.C.	53	200	100	11·2	5·8	10·1	5·0	5·0	3·0
C.P.	53	240	120	12·0	7·2	9·2	4·4	6·2	3·5
M.K.	62	250	130	14·0	8·3	8·5	4·1	6·0	3·4

dialysis with poorer biochemical control. It would seem reasonable to suggest that the currently measured parameters of nitrogen retention poorly reflect the uraemic control of the patient. Earlier work on chronic haemodialysis with once weekly treatment and PT300 cellophane gave similar biochemical control to the levels achieved currently with peritoneal dialysis but produced poor results with the production of the underdialysis syndrome. It would seem that the peritoneal membrane may be more efficient at removing larger molecular nitrogen-containing compounds. This fact will be further discussed in the section on anaemia.

Dietary and drug requirements

To control hypertension, prevent heart failure and deaths from protein intoxication or hyperkalaemia strict dietary control between dialysis is essential. All patients are instructed to eat a diet consisting of 50 g protein, 22 mEq sodium and measured fluid restriction to 500 ml. High calorific value of the diet is maintained with fats and carbohydrates aiming at an intake of 3,000 calories daily. Potassium intake is restricted to 40–50 mEq per day and all high potassium foods are excluded. During dialysis a free diet is permitted with only minimal protein restriction. No routine drug therapy is required but some patients take Resonium A ion exchange resin to control hyperkalaemia. Occasional courses of antibiotic therapy are used for intercurrent infections or infection of cannula sites. Vancomycin has proven the drug of choice for staphylococcal cannula site infections.

Blood pressure control (Table 4)

Complete resolution of malignant hypertensive pathology can be achieved with adequate dialysis and dietary restriction between dialysis (Comty, Rottka and Shaldon, 1964). All patients with malignant hypertension have expanded extracellular spaces and total body water spaces at the start of treatment in the absence of overt oedema. Progressive reduction in the sodium and water spaces by ultrafiltration during dialysis results in control of blood pressure and resolution of hypertensive pathology. Nevertheless, it is our impression that

PLATE IX

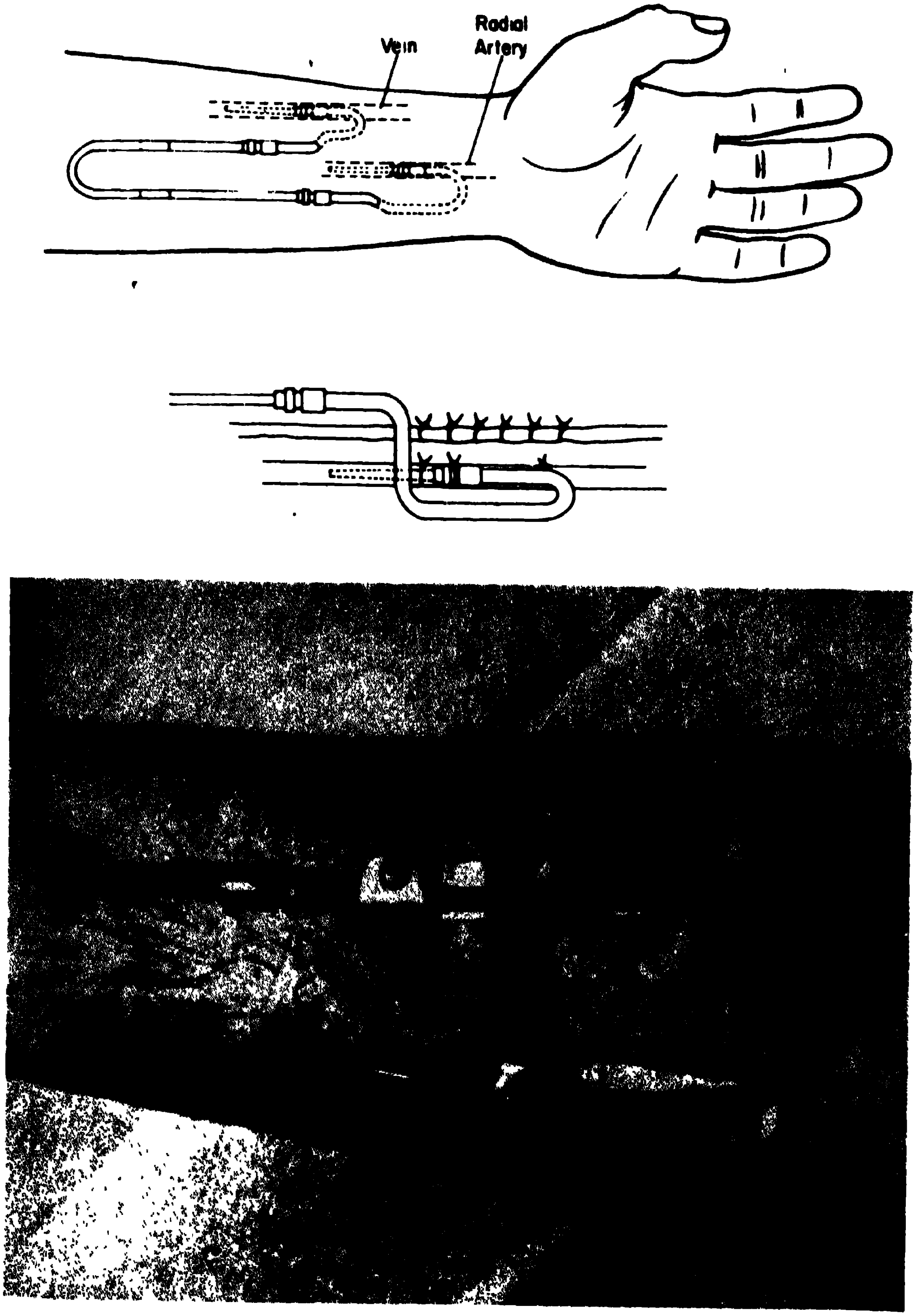


FIG. 1. Silastic-teflon cannulas implanted in the arm of a patient with cannula stabilisers in place. (Quinton *et al.*, 1962.)

blood pressure control is easier using a cuprophan membrane for haemodialysis or with peritoneal dialysis than our earlier experience with PT300 cellophane. This would suggest that the restoration of salt and water compartments to normal is not the sole factor in maintaining blood pressure control, but rather that the more efficient membranes remove an organic compound of larger molecular size which potentiates and maintains the hypertensive state. Furthermore, bilateral nephrectomy has not proven necessary in any of our patients to control hypertension.

TABLE 4. Blood pressure control

	B.P. before treatment	Average predialysis B.P.
	Haemodialysis	
E.R.	260/160 M.H.	140/90
R.J.	160/110	140/70
P.J.	180/120	150/95
R.B.	240/140 M.H.	150/95
K.H.	220/160 M.H.	140/90
O.H.	260/170 M.H.	130/70
J.B.	220/120	140/100
B.C.	160/110	140/85
J.W.	240/140 M.H.	160/100
D.M.	240/130 M.H.	160/100
W.M.	150/110	120/80
R.H.	240/140 M.H.	140/90
M.F.	240/140 M.H.	160/100
H.M.	170/110	140/80
P.W.	240/140 M.H.	160/100
A.W.	250/150 M.H.	150/90
	Peritoneal dialysis	
N.C.	220/120	140/90
C.P.	200/130	160/100
M.K.	280/110	140/90

M.H. = Malignant hypertension.

Anaemia (Table 5)

The anaemia of chronic uraemia is not entirely corrected by dialysis. In fact the average transfusion requirement remains 2 pints of blood per month for haemodialysis patients. This permits them to work a full week with a haematocrit of between 25–30 per cent (haemoglobin 8–10 g per 100 ml). Blood loss

TABLE 5

	Average predialysis haematocrit	Average transfusion requirement (RBC units) per month	Duration of treatment (months)
	Haemodialysis		
R.B.	31	0·25	16
P.J.	30	0·25	16
J.B.	28	0·25	12
R.J.	30	1·3	21
E.R.	27	2	27
K.H.	25	2	16
B.C.	28	2	11
J.W.	27	2	7
D.M.	25	2	7
M.S.	27	2	7
W.M.	28	2	5
R.H.	29	2	5
M.F.	27	2	3
H.M.	30	2	3
P.W.	30	2	3
A.W.	25	2	2
O.H.	29	3	19
	Peritoneal dialysis		
N.C.	25	0	12
C.P.	25	0·4	5
M.K.	24	0	4

in the dialyser does not account for this requirement. However, some patients may go for months without transfusion and patients receiving peritoneal dialysis virtually require no blood transfusions. This latter observation suggests that membrane research may yield the answer to anaemia in the chronic dialysis patient. Nevertheless, 80–100 per cent rehabilitation can be achieved in spite of the chronic anaemia that these patients suffer from and more intensive dialysis has reduced transfusion requirements (Shaldon, 1964). Erythropoietin has been detected in the plasma of one patient (R.B.) who although totally anuric did not require a transfusion on haemodialysis for over twelve months. It would appear that residual endocrine activity may be present in the anuric kidney, and it is of some interest that peripheral vein plasma renin levels were also high at one time in this patient.

Peripheral neuropathy (Table 6)

Chronic renal failure is now established as a cause of peripheral neuropathy (Asbury, Victor and Adams, 1963). The neuropathy mainly involves the lower limbs and is predominantly sensory although in some cases motor involvement occurs.

TABLE 6. Peripheral neuropathy

Patient	Clinical	Involvement	Conduction	Time
	sensory	motor	sensory	motor
		Haemodialysis		
R.J.	++ → O	O	R → N	R
R.B.	O	O	N	N
O.H.	++ → O	O	R → N	N
J.B.	O	O	N	N
B.C.	O	O	R	R
K.H.	+ → O	O	R	N
J.W.	+ → O	O	N	R
M.S.	+++ → +	++ → +++ → ++	R	R
W.M.	++ → +	O	R	R
P.J.	O	O	R	N
D.M.	+ → O	O	R	R
E.R.	++ → O	O	R → N	R → N
R.H.	+ → O	O	R	N
M.F.	+ → O	O	R	N
H.M.	+ → O	O	R	N
P.W.	+ → O	O	R	N
A.W.	+ → O	O	R	N
		Peritoneal dialysis		
N.C.	+ → O	O	R	N
C.P.	++ → O	O	R	N
M.K.	+ → O	O	R	N

O = no involvement
+ = minimal
++ = moderate
+++ = severe
R = reduced
N = normal

Eighty per cent of our patients had sensory symptoms and signs in their lower limbs before dialysis treatment commenced. All have improved with intensive dialysis. One patient (M.S.) had a severe motor neuropathy as well and this has now stabilized leaving him only partially disabled. Less successful results in the elimination of neuropathy have been reported from other centres (Murray, Pendras, Lindholm and Erickson, 1964). It is our belief that peripheral neuropathy can be eliminated as a disabling complication of chronic uraemia by early and intensive dialysis. Terminal patients may be left paraplegic even with successful dialysis if treatment is delayed too long. The appearance or deterioration of a peripheral neuropathy must be considered as an absolute indication for commencing dialysis therapy without delay (Konotey-Ahulu, Baillod, Comty, Heron, Shaldon and Thomas, 1965). Other indications for starting treatment include failure to control malignant hypertension with drug therapy and inability of the patient to work.

Cardiac tamponade

Pericarditis is a common complication of chronic uraemia, but a haemorrhagic tamponade is unusual. However, two patients (M.F. and H.M.) both developed severe tamponade which resolved with paracentesis and intensive dialysis. The failure to recognize tamponade may prove fatal and any patient with pericarditis should be carefully watched in the first months of treatment.

Under-dialysis syndrome

Initial reports on chronic haemodialysis (Hegstrom, Murray, Pendras, Burnell and Scribner, 1961, 1962) stressed the following complications; arthropathy, pseudogout, metastatic calcification, secondary hyperparathyroidism, failure to gain weight and progressive pigmentation. These complications are largely preventable by adequate dialysis and no longer need occur. Indeed, patient B.C. with primary gout is now symptom free and losing his tophi.

ECONOMIC ASPECTS

Chronic dialysis currently costs £2,000 per year in our hospital series and £1,000 per year in the home. As 500–1,000 suitable patients are currently dying from lack of treatment in the United Kingdom each year expansion of therapy is a matter of utmost urgency. For this reason, our unit is expanding its home dialysis programme and, to date, over six months' experience of home dialysis suggests that in selected patients, provided the correct equipment is available, home dialysis is the treatment of choice. Peritoneal dialysis costs as much as haemodialysis and in addition causes discomfort to the patient during treatment. It is unlikely that this form of treatment will supplant haemodialysis, rather it will supplement it.

CONCLUSIONS

Chronic dialysis therapy has now kept three patients alive for over five years in Seattle yet there are probably under 200 patients receiving this form of treatment throughout the world. Technical advances have demonstrated that self-dialysis and home dialysis are realities. It does not seem unreasonable to postulate that in the next decade chronic dialysis will become the 'Insulin' of the 'Chronic Nephritic'.

ACKNOWLEDGEMENTS

I would like to thank the following for allowing me to draw heavily on their own personal work, which I have presented as a member of the group: R. A. Baillod; C. Comty; M. Ilahi; F. I. D. Konotey-Ahulu; and L. Sevitt.

REFERENCES

ASBURY, A. K., VICTOR, A. and ADAMS, R. D. (1963) *Arch. Neurol.*, **8**, 413.

BAILLOD, R. A., COMTY, C., ILAHI, M., KONOTEY-AHULU, F. I. D. and SHALDON, S. (1965) *Trans. Europ. Dial. Transpl. Assn.* In press.

BARBER, N. D., NAKAMOTO, S., MCCORMACK, L. J. and KOLFF, W. J. (1963) *Trans. Amer. Soc. artif. Int. Org.*, **9**, 21.

BOEN, S. T. (1965) Personal communication.

BOEN, S. T., MION, C. M., CURTIS, F. K. and SHILIPETAR, G. (1964) *Trans. Amer. Soc. artif. int. Org.*, **10**, 409.

COLE, J. J., POLLARD, T. L. and MURRAY, J. S. (1963) *Trans. Amer. Soc. artif. int. Org.*, **9**, 67.

COMTY, C., ROTTKA, H. and SHALDON, S. (1964) *Trans. Europ. Dial. Transpl. Assn.*, **1**, 209.

CURTIS, F. K., COLE, J. J., FELLOWS, B. J., TYLER, L. L. and SCRIBNER, B. H. (1964) *Nat. Kidney Dis. Found. Ann. Conf., New York.* In press.

CURTIS, F. K., COLE, J. J., FELLOWS, B. J., TYLER, L. L. and SCRIBNER, B. H. (1965) *Trans. Amer. Soc. artif. int. Org.*, **11**, 17.

ELKINTON, J. R. (1964) *Ed. Ann. intern. Med.*, **60**, 309.

GONZALES, F. M., PABICO, R. C., BROWN, H. W., MAHER, J. F. and SCHREINER, G. E. (1963) *Trans. Amer. Soc. artif. int. Org.*, **9**, 11.

GRIMSRUD, L., COLE, J. J., LEHMAN, G. A., BABB, A. L. and SCRIBNER, B. H. (1964) *Trans. Amer. Soc. artif. int. Org.*, **10**, 107.

HEGSTROM, R. M., MURRAY, J. S., PENDRAS, J. P., BURNELL, J. M. and SCRIBNER, B. H. (1961) *Trans. Amer. Soc. artif. int. Org.*, **7**, 136.

HEGSTROM, R. M., MURRAY, J. S., PENDRAS, J. P., BURNELL, J. M. and SCRIBNER, B. H. (1962) *Trans. Amer. Soc. artif. int. Org.*, **8**, 266.

KONOTEY-AHULU, F. I. D., BAILLOD, R. A., COMTY, C., HERON, J. R., SHALDON, S. and THOMAS, P. K. (1965) *Brit. med. J.*, **ii**, 1212.

LYMAN, D. J. (1963) *Trans. Amer. Soc. artif. int. Org.*, **9**, 92.

MERRILL, J. P., SCHUPAK, E., CAMERON, E. and HAMPERS, C. I. (1964) *J. Am. med. Ass.*, **190**, 468.

MURRAY, J. S., PENDRAS, J. P., LINDHOLM, D. D. and ERICKSON, R. V. (1964) *Trans. Amer. Soc. artif. int. Org.*, **10**, 191.

PENDRAS, J. P., COLE, J. J., TU, W. H., and SCRIBNER, B. H. (1961) *Trans. Amer. Soc. artif. int. Org.*, **7**, 27.

QUINTON, W. E., DILLARD, D. H., COLE, J. J., and SCRIBNER, B. H. (1962) *Trans. Amer. Soc. artif. int. Org.*, **8**, 236.

ROBIN, E. D. (1964) *J. Am. Med. Ass.*, **189**, 624.

SCRIBNER, B. H. (1964) *Trans. Amer. Soc. artif. int. Org.*, **10**, 209.

SCRIBNER, B. H. (1965*a*) Personal communication.

SEVITT, L., COMTY, C., ROTTKA, H. and SHALDON, S. (1964) *Trans. Europ. Dial. Transpl. Assoc.*, **1**, 271.

SHALDON, S. (1964) *Trans. Amer. Soc. artif. int. Org.*, **10**, 207.

SHALDON, S., SILVA, H. and ROSEN, S. M. (1964) *Brit. med. J.*, **ii**, 411.

SHALDON, S., ROSEN, S. M. and SILVA, H. (1963) *Proc. Third Symp. Water and Elec. Metab.*, Elsevier Pub. Co., Amsterdam, p. 240.

SHERRIS, J. C., COLE, J. J. and SCRIBNER, B. H. (1961) *Trans. Amer. Soc. artif. int. Org.*, **7**, 37.

SMITH, A. H. (1965) *Clin. Sc.*, **28**, 191.

XIV

The Transport of Sodium and Potassium across Cell Membranes

I. M. GLYNN

INTRODUCTION: THE SODIUM PUMP

ALMOST all living cells contain potassium as their chief cation; yet in blood plasma and in interstitial fluid, in pond water and in sea water, sodium is much more abundant. Why cells should have evolved containing more potassium and less sodium than the fluid surrounding them is a question that has often been asked but never satisfactorily answered. We know that in nerve and muscle, and in other excitable cells, the immediate source of energy for the propagation of the impulse is the difference in concentration of sodium on the two sides of the cell membrane. Excitability is a primitive feature of living cells, and it may be that the difference in cation composition of intracellular and extracellular fluid was evolved as part of some kind of excitability mechanism. On the other hand, the difference in salt concentration may have come first and been made use of in the evolution of an excitability mechanism later. In cells living now, there are several enzymes which are activated by potassium ions and inhibited by sodium ions; but here again, it seems at least as likely that the enzymes have evolved to suit the internal medium, as that the medium has evolved to suit the enzymes. One rather plausible hypothesis to account for the difference in composition, is that the expulsion of sodium was developed as a way of overcoming the osmotic entry of water that must have presented a problem as soon as cells began to accumulate large non-penetrating molecules inside themselves (Conway, 1945). One way of dealing with this problem is to develop an inextensible tough cell wall capable of withstanding a great hydrostatic pressure, like the cell wall of plants. Another way is simply to pump anything out of the cell so that its excess concentration outside balances the osmotic pressure of the cell proteins and phosphates. And sodium would be, after all, the most abundant solute in the cell.

Whatever may have been the primary purpose of the sodium pump, in animals living now it has a variety of functions. I have already mentioned its role in providing energy for the conduction of impulses in nerve and muscle. A very similar mechanism occurs more dramatically in the electric organs of certain fish; in electric eels and torpedoes, several thousand excitable membranes are arranged in series so that the action potentials add up to produce shocks of several hundred volts. A different role is found in the cuttlefish, where the pumping of salt across a membrane controls the osmotic flow of water, which in turn determines the volume of a gas space in the cuttle bone. The volume of gas affects the density of the cuttle bone and in this way controls the buoyancy of the animal (Denton, Gilpin-Brown and Howarth, 1961).

Less bizarre but more important, salt transport in the proximal kidney tubule and the gall bladder is responsible for the concentration of urine and of bile. Further on in the kidney tubule—in the loop of Henle and in the distal convoluted tubule—sodium transport is concerned in the formation of a hyper- or hypotonic urine, in maintaining the balance of sodium and potassium in the body, and probably also in controlling the excretion of hydrogen ions. Rather more surprisingly, it now looks as though the sodium pump is involved in the uptake of glucose from the gut (Riklis and Quastel, 1958; Csáky and Thale, 1960; Csáky and Zollicoffer, 1960; Crane, Miller and Bihler, 1961; Csáky, Hartzog and Fernald, 1961; Bihler and Crane, 1962; Bihler, Hawkins and Crane, 1962; Schultz and Zalusky, 1964; Barry, Smyth and Wright, 1966), and from work particularly on ascites tumour cells it seems that it may play some part in the passage of amino acids across cell membranes (Christensen, Riggs, Fischer and Palatine, 1952; Riggs, Walker and Christensen, 1958; Gonda and Quastel, 1962; Bittner and Heinz, 1963; Kromphardt, Grobecker, Ring and Heinz, 1963; Fox, Thier, Rosenberg and Segal, 1964). Iodide uptake by thyroid, mammary gland and salivary glands seems also to be connected with cation movements (Wolff, 1960; Wolff and Maurey, 1961).

But what I want to discuss in this review is not the historical problem—why was the sodium pump evolved? Nor do I want to consider the various uses to which it has been put. Instead I want to discuss work that has been done in the past ten years or so, which tells us a little about how the sodium pump actually works.

'TRANSPORT ATPASE'

In investigating any pump, the first thing one wants to know is the source of energy. In nerve, in muscle, in the nucleated red cells of birds and reptiles, and probably in most cells, the energy appears to come from aerobic metabolism (Shanes, 1951; Maizels, 1954; Hodgkin and Keynes, 1955). In mammalian red cells the cytochrome system is lacking, and energy is provided by glycolysis (Harris, 1941; Danowski, 1941; McKee, Ormsbee, Anfinsen, Geiman and Ball, 1946; Maizels, 1951).

If in some cells pumping depends on aerobic metabolism, and in others on glycolysis, it is rather unlikely that the pump mechanism is linked directly to any of the individual stages in the glycolytic chain or in the Krebs cycle or cytochrome system. It is more likely that glycolysis and oxidative metabolism are required merely as sources of adenosine triphosphate (ATP), and that it is the ATP that provides the energy for ion transport.

There is quite a bit of indirect evidence in support of this idea, but the most direct evidence comes from experiments in which ATP has been made to enter cells lacking any other energy supply. Fig. 1, from a paper by Caldwell, Hodgkin, Keynes and Shaw (1960), shows the effect of injected ATP on the efflux of sodium from a squid giant axon that had been poisoned with cyanide. The cyanide caused a marked drop in sodium efflux, and the injection of ATP raised the efflux temporarily towards its normal value. ATP was injected twice, in two different doses, and, roughly speaking, the areas under the humps were proportional to the amounts of ATP injected, suggesting that the number of sodium ions extruded was more or less proportional to the number of molecules of ATP available.

Thirteen years ago Clarkson and Maizels (1952) showed that red cell ghosts possessed ATPase activity, and in 1957 Skou, working in Denmark, showed that a particulate fraction from minced crab nerve possessed ATPase activity which was stimulated by sodium ions and, in the presence of sodium, was further stimulated by potassium ions. Skou suggested that this ATPase might be part of the transport system.

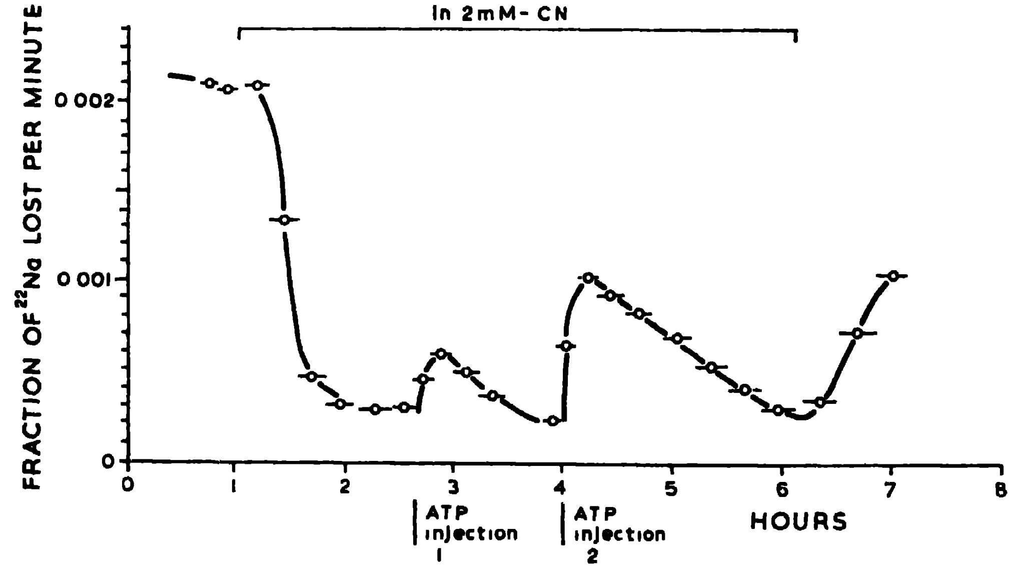


FIG. 1. The effect of injected ATP on the efflux of sodium from a squid axon that had been poisoned with cyanide. The calculated internal ATP concentration was 1·2 mM after the first injection and 6·2 mM after the second. (From Caldwell, Hodgkin, Keynes and Shaw, 1960.)

In 1958, Dr Dunham and I set out to compare the properties of the ATPase in red cell ghosts with the properties of the cation pumping system in intact red cells. Our aim was to try to decide whether there was a direct connexion between the two. Red cells had the advantage for this work that they were structurally simple—so that one knew one was dealing with cell membranes—and that a good deal was already known about the pattern of ion fluxes (see Glynn, 1957*b* for review). It turned out, from our work (Dunham and Glynn, 1961) and from similar work in America by Post, Merritt, Kinsolving and Albright (1960) at Vanderbilt and by Hoffman (1960) at

Bethesda, that there were so many resemblances between the ATPase activity and the ion transport that there was little doubt that the two processes were intimately connected.

Perhaps the most striking feature of the pump is that it is a coupled pump, in the sense that sodium ions are pumped out only if potassium ions are available outside the cell to be pumped in. In red cells, the first suggestion that the pump was coupled came from Harris and Maizels (1951), who noticed that cells suspended in a solution poor in potassium seemed to pump out less sodium. More convincing evidence for coupling is shown in Fig. 2. The part of the sodium efflux that requires glucose—and is therefore presumably active—is reduced as the external potassium concentration is lowered, and vanishes when potassium is absent. Furthermore, the curve relating active sodium efflux to external potassium concentration resembles quite closely the curve relating active potassium influx to external potassium concentration. If the pump pumps sodium out only when potassium is available to be pumped in, then we might expect an ATPase that is connected with the pump to show

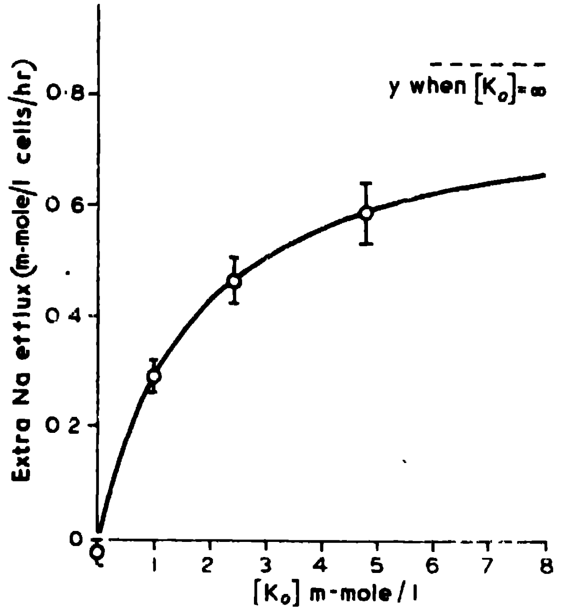


FIG. 2. The effect of external potassium concentration on the extra efflux of sodium which occurs in the presence of glucose. The experiment was performed on human red cells loaded with radiosodium. The continuous line represents the equation $y = 0{\cdot}8\,[K_o]/(1{\cdot}8 + [K_o])$, which is similar to the expression relating active potassium influx to external potassium concentration. (From Glynn, 1956.)

its activity only when both sodium and potassium are present. And this is what happens. Table 1 shows the amount of inorganic phosphate liberated from ATP in one hour by the red cell ghosts from one litre of intact cells. There was considerable activity in the absence of both sodium and potassium; when sodium or potassium was added alone there was either no increase or at most a very small increase in activity; when they were added together the activity was roughly doubled.

TABLE 1. The effects of Na^+, K^+ and cardiac glycosides on the ATPase activity of red cell ghosts. Mg^{2+} was present in all samples together with sufficient Tris to make the solutions isotonic. (Dunham and Glynn, 1961.)

	[Na] (mM)	[K] (mM)	Activity (m-mole phosphate/l cells/hr)
Expt. 1	0	0	0·75
			0·78
	0	16	0·82
			0·84
	16	0	0·73
			0·83
	16	16	1·34
			1·37
In the presence of ouabain (10^{-4} g/ml)	16	16	0·78
			0·85
Expt. 2	0	0	0·57
	0	16	0·64
	64	0	0·55
	64	16	1·33
In the presence of scillaren A ($1{\cdot}7 \times 10^{-5}$ g/ml)	64	16	0·58

A second feature of the pump is that it is inhibited by low concentrations of cardiac glycosides—substances like digitalis and ouabain—(Schatzmann, 1953; see also review by Glynn, 1964); and we found that stimulation of the ATPase by sodium plus potassium could be abolished by cardiac glycosides at similarly low concentrations (Table 1). So it looked as though the ATPase activity of the red cell membrane consisted of two parts: one part which did not require sodium or potassium and which was not sensitive to cardiac glycosides; and another part which required the presence of both sodium and potassium and which was inhibited by cardiac glycosides at the concentration at which they inhibited the pump.

Although we cannot be certain, it is likely that these two parts of the ATPase represent two separate enzymes; it is of course the second part that we are interested in. One reason for thinking that there are two different enzymes is that calcium ions have opposite effects. The part of the ATPase which requires sodium plus potassium is strongly inhibited by calcium ions; the other part is stimulated (Dunham and Glynn, 1961). It is convenient to have a quick way of referring to the part of the ATPase that requires sodium plus potassium and is sensitive to cardiac glycosides, and I shall therefore use the phrase 'transport ATPase', realizing that at this stage I am begging the question in doing so.

TABLE 2. The effects of Na^+, K^+ and ouabain on the ATPase activity of a preparation of *Torpedo* electric organ. (Data from Glynn, Slayman, Eichberg and Dawson, 1965.)

	[Na] mM	[K] mM	Activity (as per cent activity in complete medium)
	106	0	1·9
	0	20	0·9
	106	20	100
In the presence of ouabain [10^{-4}] g/ml	106	20	1·4

In red cells, the presence of the uninteresting ATPase complicates the picture. Table 2 shows the ATPase activity of a preparation of the electric organ of a torpedo. In this preparation the omission of sodium or potassium or the addition of ouabain led to the loss of about 98 per cent of the activity—in other words, nearly all the activity could be attributed to 'transport ATPase'.

A third feature of the pump is that the inhibitory effect of cardiac glycosides is much reduced by certain modifications to the glycoside molecules (Kahn and Acheson, 1955; Glynn, 1957*a*, 1964). The same modifications similarly reduced the inhibitory effect on the 'transport ATPase' (Dunham and Glynn, 1961).

A fourth and rather peculiar feature of the pump is that inhibition by low concentrations of a cardiac glycoside can be prevented or reduced by raising the potassium concentration outside the cell (Glynn, 1957*a*). Table 3 shows that potassium has a similar effect on the inhibition of the 'transport ATPase'. Again there is a close resemblance between the properties of the pump in intact cells and the properties of the (Na + K)—activated ATPase in red cell ghosts.

TABLE 3. The effect of potassium concentration on the inhibition by ouabain of the glycoside-sensitive ATPase activity of red cell ghosts. (Glynn, 1957*a*.)

[K] (mM)	Total glycoside-sensitive activity (m-mole phosphate/l cells/hr)	Activity in the presence of ouabain 5×10^{-8} g/ml (m-mole phosphate/l cells/hr)	Inhibition (%)
0·25	0·188	0·027	86
0·5	0·364	0·134	63
1	0·619	0·290	53
2	0·778	0·498	36
4	0·855	0·686	20
8	0·926	0·782	15
16	1·02	0·823	19
32	1·04	0·964	7

If the ATPase that is activated by sodium plus potassium is really part of the pump mechanism, then it is a reasonable guess that the activating ions are the ions that are being pumped. And since sodium is pumped out of the cell and potassium is pumped into it, we can predict that activation of the ATPase should occur only when potassium ions are present at the outer surface of the cell membrane and sodium ions are present at the inner surface. The experiments that I have described give no direct information on this point because they were all done with ghosts that were so leaky that the concentrations of all ions in the solution must have been nearly the same on both sides of the membrane. To test the effects of particular ions on one or other side of the membrane, it is necessary to use ghosts with relatively impermeable membranes. These can be obtained by a process, sometimes called reversible haemolysis, discovered independently by Székely, Mányai and Straub (1952) in Hungary, and by Teorell (1952) in America. If you lyse red cells in water or in hypotonic solutions, the cells swell up and become spherical. Further swelling then stretches the membranes, which become leaky and allow the contents of the cells to diffuse out. Most of the cells do not show any gross rupture. It has been known for many years that if sufficient salt is added to the suspension of lysed cells to restore isotonicity, the cells regain their disc shape; and what Straub and his collaborators found was that under certain conditions a proportion of the cells also regain, to a large extent, their impermeability to cations. The resealing can be improved by incubating the cells at 37° C. after the addition of salt (Hoffman, Tosteson and Whittam, 1960). At the moment of lysis the cells are extremely leaky, so that any reasonably small molecules in the lysing solution will enter at that moment and will be trapped when the cells are resealed. This technique may be used to prepare what Straub originally called reversed cells—'revertierten Zellen'—but what have come to be called by most workers in this field 'reconstituted ghosts'. You can get these ghosts containing any ions you like, including ATP (Gardos, 1954), and you can then put them in any desired solution and measure the ATPase activity. Table 4 shows the results of an experiment of this kind. With potassium inside the cells and with sodium outside, there was very little glycoside-sensitive splitting of ATP, whereas with potassium outside and mostly sodium inside there was rapid splitting. Since the cells inevitably contained a trace of potassium—the remains of the potassium that was in the cells before they were lysed—a third kind of cell suspension was prepared in which the cells contained both sodium and potassium but were suspended in a medium containing only sodium. These cells, too, showed very little glycoside-sensitive ATPase activity. Now, since the very small amounts of activity in the 'unfavourable' conditions can reasonably be attributed to a slight leakage of sodium inwards and of

TABLE 4. Activation of the glycoside-sensitive ATPase of 'reconstituted' red cell ghosts by external potassium and internal sodium. (Glynn, 1962.)

Cells		Incubation medium		Glycoside-sensitive ATPase activity
[K]	[Na]	[K]	[Na]	m-mole phosphate/l cells/hr
9	0	0	25	0·048
1	25	8	0	0·477
9	25	0	25	0·037

potassium outwards, it seems that, taken together, the results show that our prediction was correct; that is to say, glycoside-sensitive ATPase activity is found only when potassium is present outside and sodium is present inside. Results similar to those shown in Table 4 have also been obtained by Whittam (1962), at Oxford.

THE STOICHIOMETRY OF THE PUMP

Having got this far we can go one step further and use our reconstituted ghosts to get a figure for the stoichiometry of the pump—the ratio of sodium ions pumped to molecules of ATP split. To do this we need to compare the rate at which ATP is split by the glycoside-sensitive ATPase with the rate at which sodium is lost by the glycoside-sensitive pathway—both measurements being made at the same time, under the same conditions and on cells from the same batch. Table 5 shows the results of an experiment of this kind. To get a figure for the glycoside-sensitive loss of sodium, it is not good enough simply to take the difference between the sodium loss in the presence and in the absence of cardiac glycoside; in the presence of glycoside the cells will contain slightly more sodium than they would have contained if the glycoside had not been there, and so the passive loss will be slightly greater. To correct for this, I have assumed that the passive loss is proportional to the internal sodium concentration—this is probably correct, but in any case the error introduced will be very small as the correction itself is

TABLE 5. Stoichiometry of the sodium pump. (Pumping downhill.) (Glynn, 1962.)

Na loss m-equiv/l cells/hr	control	6·01 ± 0·10	difference 3·43 ± 0·16	corrected for change in passive efflux 3·65 ± 0·17
	ouabain	2·58 ± 0·13		
ATP hydrolysed m-mole/l cells/hr	control	2·20 ± 0·02	difference 1·11 ± 0·03	
	ouabain	1·09 ± 0·02		

$$\text{Ratio}\ \frac{\text{ouabain-sensitive efflux of Na}}{\text{ouabain-sensitive hydrolysis of ATP}} = 3{\cdot}29 \pm 0{\cdot}18$$

quite small. Comparing the corrected figure for glycoside-sensitive sodium efflux with the figure for glycoside-sensitive ATPase activity we get a ratio of 3·29 ± 0·18—that is, not significantly different from 3.

In this experiment the cells were pumping sodium into a sodium-free solution. Using a rather different technique, Sen and Post (1964) found a similar figure for the stoichiometry when the cells were pumping sodium into a solution containing sodium at the same concentration as was present internally. So far, for various technical reasons, no one has succeeded in doing the experiment under conditions in which the pumping is uphill.[1]

Indirect evidence that the stoichiometry is the same when sodium is being pumped uphill comes from work on frog skin by Zerahn (1956) and by Leaf and Renshaw (1957). These authors compared the amount of sodium transported across the skin with the extra amount of oxygen used when transport took place, and found that about 18 sodium ions were transported per molecule of oxygen; the ratio seemed to be much the same whether the pumping was uphill or not. Assuming each atom of oxygen gives rise to 3 energy-rich phosphate bonds, a figure of 18 sodium ions per molecule of oxygen is equivalent to 3 sodium ions per molecule of ATP.

One rather intriguing point is that the figure of 3 sodium ions pumped per molecule of ATP split fits in with a curious observation made by Keynes and Swan (1959) on frog muscles that had been depleted of sodium so that the internal concentration was low. Under these conditions, the efflux of sodium varied with the cube of the internal sodium concentration. One way of explaining both this cubic relation and the 3 : 1 stoichiometry is to suppose that the pump involves a carrier which moves across the membrane every time a molecule of ATP is split, and that the carrier has three sites all of which must be occupied by sodium ions if the mechanism is to work.

---

[1] Experiments have recently been reported on red cells (Whittam, R. and Ager, Margaret E. (1965) *Biochem. J.*, 97, 214) and on crab nerves (Baker, P. F. (1965) *J. Physiol.*, 180, 383) suggesting that a stoichiometric ratio of about 3 Na: 1 ATP also holds for uphill sodium transport.

DISTRIBUTION OF 'TRANSPORT ATPASE' ACTIVITY

So far, I have been talking mainly about red cell ATPase. If the transport of sodium and potassium is necessarily associated with ATPase activity, one ought to find ATPases with similar properties in almost any tissue that one cares to look at. Within the last few years very active ATPases requiring sodium plus potassium have been described in brain and kidney preparations (Deul and McIlwain, 1961; Järnefelt, 1961, 1962*a*; Whittam and Wheeler, 1961; Aldridge, 1962; Cummins and Hydén, 1962; Schwartz, Bachelard and McIlwain, 1962; Skou, 1962; Somogyi and Vincze, 1962; Wheeler and Whittam, 1962; Yoshida and Fujisawa, 1962; Bonting and Caravaggio, 1963; Jöbsis and Vreman, 1963; Kinsolving, Post and Beaver, 1963; Landon and Norris, 1963; Van Groningen and Slater, 1963; Rodnight and Lavin, 1964); and Dr. Bonting and his colleagues have shown them to be present to a greater or lesser extent in nearly every tissue in the cat, and in a number of tissues in various other animals (Bonting and Caravaggio, 1962, 1963; Bonting, Caravaggio and Canady, 1964; Bonting, Caravaggio, Canady and Hawkins, 1964; Bonting, Caravaggio and Hawkins, 1962, 1963). These authors also showed that there was a fairly good correlation between the amount of cation pumping going on in a tissue and the activity of the ATPase in the tissue that required the presence of sodium plus potassium. This correlation held good over a very wide range, from the actively pumping electric organs, at one end of the scale, to the sluggish mammalian red cells at the other.

A similar correlation was found by Tosteson, Moulton and Blaustein (1960), who compared cation transport and (Na + K)-dependent ATPase activity in the red cells of different breeds of sheep. In foetal sheep, of any breed, the red cells pump sodium and potassium vigorously. Adult sheep are of two kinds: one kind has cells very rich in potassium and with a vigorous Na-K exchange pump; the other kind has a rather feeble pump and cells correspondingly poorer in potassium. It turned out that the cells with the vigorous pump were rich in glycoside-sensitive ATPase activity and the cells with a feeble pump had much less. Similar comparisons have been made between the potassium-rich red cells found in guinea-pig and man, and the potassium-poor red cells of cats. Again the cells poor in potassium had very much less glycoside-sensitive ATPase activity (Greeff, Grobecker and Piechowski, 1964).

INTERMEDIATES IN THE TRANSPORT SYSTEM

I think that what has been said justifies the term 'transport ATPase'. Now what in fact is this ATPase? What happens to the ATP before it ends up as inorganic phosphate and ADP?

The most direct approach to this problem has been to try to find phosphorylated intermediates. That such intermediates exist seems likely on general grounds and has been demonstrated in two ways. Skou (1960) showed that a microsomal preparation, from crab nerve, which possessed (Na + K)-dependent ATPase activity catalysed an ATP-ADP exchange reaction; that is to say, after incubation of the preparation with ATP in the presence of ^{32}P-labelled ADP, radioactivity could be found in the ATP. The rate of the exchange reaction could be reduced by adding sodium and potassium together, and the reduction in rate could be prevented by adding ouabain. Skou therefore suggested that the first step in the splitting of ATP was a reversible reaction between ATP and an enzyme in the membrane to form a phosphorylated enzyme together with ADP.

The other evidence that the transport ATPase system involves phosphorylated intermediates comes from experiments in which membrane preparations possessing (Na + K)-dependent ATPase activity were exposed briefly to ^{32}P-labelled ATP and then examined for incorporated radioactivity. Experiments of this kind have been done on kidney preparations by Charnock and Post (1963), Charnock, Rosenthal and Post (1963), Post, Sen and Rosenthal (1965) and Whittam, Wheeler and Blake (1964); on a brain preparation by Rodnight and Lavin (1964); and on preparations of electric organ by Albers, Fahn and Koval (1963) and by Mrs. Slayman and me (unpublished). (See also related work on phosphoprotein turnover in liver slices and red cells by Judah and Ahmed (1962) and Judah, Ahmed and McLean (1962).) There seems to be general agreement that in the presence of sodium, some of the ^{32}P is incorporated into the preparation; if potassium is present as well as sodium then less ^{32}P is incorporated, but the effect of potassium can be prevented by ouabain. Post, Sen and Rosenthal (1965) have also shown that potassium accelerates the loss of ^{32}P already incorporated. This all fits in with the hypothesis that sodium is required for a phosphorylation step and potassium is required for dephosphorylation.

The idea that the transport of potassium is associated with dephosphorylation is also supported by evidence of quite a different kind. Very recently Dr. Garrahan and I (1965) have found that under certain conditions the transport system in red cells can be uncoupled, so that in the absence of potassium there is simply a catalysed exchange of sodium ions across the membrane. For this to take place ATP must be present but the rate of splitting is very low.

Now, unfortunately, it is one thing to demonstrate that intermediates exist; it is quite another to find out what they are. For many years, although there is no direct evidence for such a theory, it has been fashionable to explain the sodium pump in terms of a circulation of lipid-soluble carriers specific for sodium or potassium ions, and it has often been thought that these carriers might be phospholipids. In 1960 Hokin and Hokin reported some interesting experiments on slices of albatross salt gland—a gland which secretes a very concentrated salt solution—and found that when the gland was stimulated to secrete by the addition of acetyl choline, the rate of incorporation of inorganic ^{32}P into phosphatidic acid was greatly increased. They extended this observation in studies of red cell ghosts and of brain microsomal preparations, and they also found evidence implicating monophosphoinositide (Hokin, Hokin and Mathison, 1963). On the basis of these findings they have built up fairly elaborate theories of salt transport, which can be found conveniently summarized in a recent review (Hokin and Hokin, 1963).

A third phospholipid suggested as a possible intermediate in 'transport ATPase' activity is triphosphoinositide, a compound which has been shown to undergo rapid dephosphorylation and rephosphorylation in brain (Wagner, Lissau, Hölzl and Horhammer, 1961; Brockerhoff and Ballou, 1962; Ellis and

Hawthorne, 1962; Dawson and Thompson, 1964; Thompson and Dawson, 1964).

I have not the space to review the Hokins' theories, or the evidence on which they are based, in any detail, but there are two general points of criticism that I should like to make.

The first point is that in preparations containing whole cells, an alteration in the quantity or in the rate of labelling of a substance on the addition of potassium or of a cardiac glycoside is not good evidence that the substance is involved in ion transport. The addition of potassium or of glycoside, by allowing or preventing active transport, may alter the ratio of ATP to ADP inside the cell and hence the whole metabolic balance.

The second point is this. About nine years ago, I looked at the effects of very low concentrations of a cardiac glycoside on the entry of potassium into red cells, and simultaneously measured the amount of glycoside that was taken up by the cells (Glynn, 1957*a*). The results suggested that the number of sites on each red cell at which active transport occurred was not more than 1,000, or something of that order. Assuming that three sodium ions are transported for each ATP split, then from the known sodium efflux one can calculate that the turnover number at each active site cannot be less than about 1,000/minute. It therefore seems likely that in any preparation of 'transport ATPase' any phosphorylated intermediate will become fully labelled within an immeasurably short time: substances in which the labelling increases over several minutes are not likely to be the intermediate one is looking for. Quite apart from this general criticism, in brain microscomes, Järnefelt (1962*b*) has made a direct comparison of the rate of splitting of γ ^{32}P-ATP with the rate of incorporation of ^{32}P into the total lipid fraction. He concluded that the incorporation was much too slow to represent the formation of an intermediate in the ATPase reaction. On the other hand, this kind of argument can never exclude the possibility that a phospholipid is an intermediate, because it is always possible that the rate of labelling of different fractions of the phospholipid is different; part of the phospholipid might be turning over much more quickly.

In view of the basic attractiveness of the idea that phospholipids might be intermediates in the ATPase system, and the fact that, although there were difficulties, there was quite a bit of evidence of one kind or another in its favour, Dr. Dawson, Dr. Eichberg, Mrs. Slayman and I thought it would be a good idea to put the matter to a more direct experimental test.

We prepared fractions rich in 'transport ATPase' activity from the electric organs of electric eels and torpedoes, and exposed them briefly to γ^{32}P-ATP of high specific activity. After 30 seconds the reaction was stopped with chloroform-methanol, and the phospholipids were extracted from the preparation, separated and assayed for radioactivity. As we were not able to predict the conditions that might favour the accumulation of any phosphorylated intermediate, we made the exposure under several different conditions. In the first place we did it under conditions that allowed hydrolysis to take place, i.e. in the presence of both sodium and potassium. Secondly we did it with both sodium and potassium present and also sufficient ouabain to inhibit the ATPase activity. Thirdly, we did it in the absence of sodium, and fourthly in the absence of potassium. In this way we hoped that, by interfering with the reaction mechanism at different points, we would improve the chances of trapping an intermediate. In fact, we found that whatever the conditions there was no significant labelling of any of the lipids.

Before concluding that no phospholipid serves as a phosphorylated intermediate in the transport ATPase system, we must consider two questions.

First, were the enzyme preparations exposed to the labelled ATP long enough for any hypothetical intermediate to become labelled? The preparations were exposed for 30 seconds. During that time between 10 and 20 per cent of the ATP was hydrolysed; the time necessary for labelling must therefore be much less than 30 seconds.

The second question is: is the amount of intermediate that one can reasonably expect to have been present enough for the labelling to have been detectable if it occurred? Theoretically, of course, if the turnover rate were fast enough the amount of intermediate could be extremely small, and experiments like those we did could not exclude the possibility that a small fraction of the phospholipid took part. But that experiment could put an upper limit on the size of the participating fraction. For phosphatidic acid it works out that not more than one part in 10,000 of the total amount present could have been involved. Another way of approaching the problem is to suppose that the turnover rate is the same as we think it might be in red cells, and then to calculate how much incorporation of ^{32}P might reasonably have been expected. If the result of this calculation is compared with the observed incorporation, it turns out that the observed incorporation was about a million times too small.

We therefore feel fairly confident that none of the phospholipids examined serves as a phosphorylated intermediate in the pump system.

ION TRANSPORT AND OXIDATIVE PHOSPHORYLATION

I should like to finish by mentioning briefly some experiments of quite a different kind that suggest an interesting possibility, though it may be a red herring. There are obvious resemblances between the 'transport ATPase' that I have been discussing and the ATPase activity of mitochondria. The normal role of mitochondria is to carry out oxidative phosphorylation—the coupling of oxidative processes with the synthesis of ATP—but it is now well known that the system responsible for oxidative phosphorylation can be made to run backwards, so that the splitting of ATP is accompanied by the passage of electrons the 'wrong way' down the oxidation chain (Chance, 1961; Klingenberg and Schollmeyer, 1961). From time to time, it has been suggested that this process might somehow be coupled to ion transport.

Experiments to test whether a similar hypothesis might explain ion transport across cell membranes provided no evidence implicating oxidation-reduction processes, but did show that the 'transport ATPase' was inhibited to a considerable extent by oligomycin, an antibiotic formerly thought to be a specific inhibitor of oxidative phosphorylation (Glynn, 1963; Järnefelt, 1962*a*; Jöbsis and Vreman, 1963; Van Groningen and Slater, 1963). In red cells I found that oligomycin partly inhibited the active transport of sodium, and since red cells do not carry out oxidative phosphorylation, and indeed lack the enzymes for doing so, the inhibition could not have been secondary to inhibition of the oxidative phosphorylation system. This conclusion was questioned by Van Rossum (1964), who claimed that oligomycin inhibited sodium transport in liver slices, which depend on oxidative metabolism for their energy, but not in red cells, which get their energy from glycolysis. Recently, however, Whittam, Wheeler and Blake (1964) at Oxford, have confirmed that oligomycin does in fact inhibit cation transport in red cells—and can do so completely—and they have also got some interesting results with a membrane preparation from kidney cortex, in which they find that oligomycin has effects on the incorporation of ^{32}P from terminally labelled ATP.

Now oligomycin is generally believed to inhibit oxidative phosphorylation at the point where inorganic phosphate enters the system, and the question obviously arises: is the hypothetical phosphorylated intermediate of 'transport ATPase' the same as the hypothetical phosphorylated intermediate of oxidative phosphorylation? Added point to this question is given by the recent findings that mitochondria can accumulate calcium and manganese in a way that resembles true active transport, and that when this accumulation is energized by ATP it can be inhibited by oligomycin (Brierly, Murer, Bachmann and Green, 1963; Chappell, Greville, and Bicknell, 1962; Chappell, Cohn and Greville, 1963; Lehninger, 1964).

Of course, it is no answer to one unsolved problem, to suggest that it has the same solution as another unsolved problem—but at least it may reduce the number of problems.

REFERENCES

Albers, R. W., Fahn, S. and Koval, G. J. (1963) *Proc. Nat. Acad. Sci. Wash.*, 50, 474.

Aldridge, W. N. (1962) *Biochem. J.*, 83, 527.

Barry, R. J. C., Smyth, D. H. and Wright, E. M. (1966) *J. Physiol.* In press.

Bihler, I. and Crane, R. K. (1962) *Biochem. biophys. Acta*, 59, 78.

BIHLER, I., HAWKINS, K. A. and CRANE, R. K. (1962) *Biochem. biophys. Acta*, **59**, 94.

BITTNER, J. and HEINZ, E. (1963) *Biochim. biophys. Acta*, **74**, 392.

BONTING, S. L. and CARAVAGGIO, L. L. (1962) *Nature, Lond.*, **194**, 1180.

BONTING, S. L. and CARAVAGGIO, L. L. (1963) *Arch. Biochem. Biophys.*, **101**, 37.

BONTING, S. L., CARAVAGGIO, L. L. and CANADY, M. R. (1964) *Exp. Eye Res.*, **3**, 47.

BONTING, S. L., CARAVAGGIO, L. L., CANADY, M. R. and HAWKINS, N. M. (1964) *Arch. Biochem. Biophys.*, **106**, 49.

BONTING, S. L., CARAVAGGIO, L. L. and HAWKINS, N. M. (1962) *Arch. Biochem. Biophys.*, **98**, 413.

BONTING, S. L., CARAVAGGIO, L. L. and HAWKINS, N. M. (1963) *Arch. Biochem. Biophys.*, **101**, 47.

BRIERLEY, G., MURER, E., BACHMANN, E. and GREEN, D. E. (1963) *J. biol. Chem.*, **238**, 3482.

BROCKERHOFF, H. and BALLOU, C. E. (1962) *J. biol. Chem.*, **237**, 49.

CALDWELL, P. C., HODGKIN, A. L., KEYNES, R. D. and SHAW, T. I. (1960) *J. Physiol.*, **152**, 561.

CHANCE, B. (1961) *Nature, Lond.*, **189**, 719.

CHAPPELL, J.B., COHN, M. and GREVILLE, G. D. (1963) In *Energy-linked functions of mitochondria*, ed. B. Chance, Academic Press: N.Y., p. 219.

CHAPPELL, J. B., GREVILLE, G. D. and BICKNELL, K. E. (1962) *Biochem. J.*, **84**, 61 P.

CHARNOCK, J. S. and POST, R. L. (1963) Evidence of the mechanism of ouabain inhibition of cation activated adenosinetriphosphatase. *Nature, Lond.*, **199**, 910.

CHARNOCK, J. S., ROSENTHAL, A. S. and POST, R. L. (1963) *Aust. J. exp. Biol. med. Sci.*, **41**, 675.

CHRISTENSEN, H. N., RIGGS, T. R., FISCHER, H. and PALATINE, I. M. (1952) *J. biol. Chem.*, **198**, 1.

CLARKSON, E. M. and MAIZELS, M. (1952) *J. Physiol.*, **116**, 112.

CONWAY, E. J. (1945) *Biol. Rev.*, **20**, 56.

CRANE, R. K., MILLER, D. and BIHLER, I. (1961) In *Symposium on Membrane Transport and Metabolism*, ed. Kleinzeller, A. and Kotyk, A., Academic Press, London and N.Y., p. 439.

CSÁKY, T. Z., HARTZOG, H. G., III and FERNALD, G. W. (1961) *Amer. J. Physiol.*, **200**, 459.

CSÁKY, T. Z. and THALE, M. (1960) *J. Physiol.*, **151**, 59.

CSÁKY, T. Z. and ZOLLICOFFER, L. (1960) *Amer. J. Physiol.*, **198**, 1056.

CUMMINS, J. and HYDÉN, H. (1962) *Biochim., biophys. Acta*, **60**, 271.

DANOWSKI, T. S. (1941) *J. biol. Chem.*, **139**, 693.

DAWSON, R. M. C. and THOMPSON, W. (1964) *Biochem. J.*, **91**, 244.

DENTON, E. J., GILPIN-BROWN, J. B. and HOWARTH, J. V. (1961) *J. mar. biol. Ass. U.K.*, **41**, 351.

DEUL, D. H. and MCILWAIN, H. (1961) *J. Neurochem.*, **8**, 246.

DUNHAM, E. T. and GLYNN, I. M. (1961) *J. Physiol.*, **156**, 274.

Ellis, R. B. and Hawthorne, J. N. (1962) *Acta neurol. scand.*, **38** (Suppl. 1), 33.

Fox, M., Thier, S., Rosenberg, L. and Segal, S. (1964) *Biochim. biophys. Acta*, **79**, 167.

Gardos, G. (1954) *Acta physiol. Acad. Sci. Hung.*, **6**, 191.

Garrahan, P. J. and Glynn, I. M. (1965) *Nature, Lond.*, **207**, 1098.

Glynn, I. M. (1956) *J. Physiol.*, **134**, 278.

Glynn, I. M. (1957*a*) *J. Physiol.*, **136**, 148.

Glynn, I. M. (1957*b*) *Progr. in Biophys. and Biophys. Chem.*, **8**, 241.

Glynn, I. M. (1962) *J. Physiol.*, **160**, 18 P.

Glynn, I. M. (1963) *J. Physiol.*, **169**, 452.

Glynn, I. M. (1964) *Pharmacol. Rev.*, **16**, 381.

Glynn, I. M., Slayman, C. W., Eichberg, J. and Dawson, R. M. C. (1965) *Biochem. J.*, **94**, 692.

Gonda, O. and Quastel, J. H. (1962) *Biochem. J.*, **84**, 394.

Greeff, K., Grobecker, H. and Piechowski, U. (1964) *Naturwiss.*, **51**, 42.

Harris, E. J. and Maizels, M. (1951) *J. Physiol.*, **113**, 506.

Harris, J. E. (1941) *J. Biol. Chem.*, **141**, 579.

Hodgkin, A. L. and Keynes, R. D. (1955) *J. Physiol.*, **128**, 28.

Hoffman, J. F. (1960) *Fed. Proc.*, **19**, 127.

Hoffman, J. F., Tosteson, D. C. and Whittam, R. (1960) *Nature, Lond.*, **185**, 186.

Hokin, L. E. and Hokin, M. R. (1960) *J. gen. Physiol.*, **44**, 61.

Hokin, L. E. and Hokin, M. R. (1963) *Ann. Rev. Biochem.*, **32**, 553.

Hokin, L. E., Hokin, M. R. and Mathison, D. (1963) *Biochim. biophys. Acta*, **67**, 485.

Járnefelt, J. (1961) *Biochim. biophys. Acta*, **48**, 104.

Járnefelt, J. (1962*a*) *Biochim. biophys. Acta*, **59**, 643.

Járnefelt, J. (1962*b*) *Biochim. biophys. Acta*, **59**, 655.

Jöbsis, F. F. and Vreman, H. J. (1963) *Biochim. biophys. Acta*, **73**, 346.

Judah, J. D. and Ahmed, K. (1962) *Nature, Lond.*, **194**, 382.

Judah, J. D., Ahmed, K. and McLean, A. E. M. (1962) *Biochim. biophys. Acta*, **65**, 472.

Kahn, J. B. Jr. and Acheson, G. H. (1955) *J. Pharmacol.*, **115**, 305.

Keynes, R. D. and Swan, R. C. (1959) *J. Physiol.*, **147**, 591.

Kinsolving, C. R., Post, R. L. and Beaver, D. L. (1963) *Biochim. biophys. Acta*, **62**, 85.

Klingenberg, M. and Schollmeyer, P. (1961) *Biochem. biophys. Res. Communs.*, **4**, 323.

Kromphardt, H., Grobecker, H., Ring, K. and Heinz, E. (1963) *Biochim. biophys. Acta*, **74**, 549.

Landon, E. J. and Norris, J. L. (1963) *Biochim. biophys. Acta*, **71**, 266.

Leaf, A. and Renshaw, A. (1957) *Biochem. J.*, **65**, 82.

Lehninger, A. L. (1964) *The Mitochondrion.* Chapter 8. Benjamin: N.Y.

McKee, R. W., Ormsbee, R. A., Anfinsen, C. B., Geiman, Q. M. and Ball, E. G. (1946) *J. exptl. Med.*, **84**, 569.

Maizels, M. (1951) *J. Physiol.*, **112**, 59.

MAIZELS, M. (1954) *J. Physiol.*, **125**, 263.

POST, R. L., MERRITT, C. R., KINSOLVING, C. R. and ALBRIGHT, C. D. (1960) *J. biol. Chem.*, **235**, 1796.

POST, R. L., SEN, A. K. and ROSENTHAL, A. S. (1965) *J. biol. Chem.*, **240**, 1437.

RIGGS, T. R., WALKER, L. M. and CHRISTENSEN, H. N. (1958) *J. biol. Chem.*, **233**, 1479.

RIKLIS, E. and QUASTEL, H. H. (1958) *J. Biochem. Physiol., Canad.*, **36**, 347.

RODNIGHT, R. and LAVIN, B. A. (1964) *Biochem. J.*, **91**, 24 P.

SCHATZMANN, H. J. (1953) *Helv. physiol. et pharmacol. Acta*, **11**, 346.

SCHULTZ, S. G. and ZALUSKY, R. (1964) *J. gen. Physiol.*, **47**, 1043.

SCHWARTZ, A., BACHELARD, H. S. and MCILWAIN, H. (1962) *Biochem. J.*, **84**, 626.

SEN, A. K. and POST, R. L. (1964) *J. biol. Chem.*, **239**, 345.

SHANES, A. M. (1951) *J. gen. Physiol.*, **34**, 795.

SKOU, J. C. (1957) *Biochim. biophys. Acta*, **23**, 394.

SKOU, J. C. (1960) *Biochim. biophys. Acta*, **42**, 6.

SKOU, J. C. (1962) *Biochim. biophys. Acta*, **58**, 314.

SOMOGYI, J. and VINCZE, I. (1962) *Acta physiol. Acad. Sci. Hung.*, **21**, 29.

SZÉKELY, M., MÁNYAI, S. and STRAUB, F. B. (1952) *Acta physiol. Acad. Sci. Hung.*, **3**, 571.

TEORELL, T. (1952) *J. gen. Physiol.*, **35**, 669.

THOMPSON, W. and DAWSON, R. M. C. (1964) *Biochem. J.*, **91**, 233.

TOSTESON, D. C., MOULTON, R. H. and BLAUSTEIN, M. (1960) *Fed Proc.*, **19**, 128.

VAN GRONINGEN, H. E. M. and SLATER, E. C. (1963) *Biochim. biophys. Acta*, **73**, 527.

VAN ROSSUM, G. D. V. (1964) *Biochim. biophys. Acta*, **82**, 556.

WAGNER, H., LISSAU, A., HÖLZL, J. AND HORHAMMER, L. (1961) *Biochem. Z.*, **335**, 312.

WHEELER, K. P. and WHITTAM, R. (1962) *Biochem. J.*, **85**, 495.

WHITTAM, R. (1962) *Biochem. J.*, **84**, 110.

WHITTAM, R. and WHEELER, K. P. (1961) *Biochim. biophys. Acta*, **51**, 622.

WHITTAM, R., WHEELER, K. P. and BLAKE, A. (1964) *Nature, Lond.*, **203**, 720.

WOLFF, J. (1960) *Biochim. biophys. Acta*, **38**, 316.

WOLFF, J. and MAUREY, J. R. (1961) *Biochim. biophys. Acta*, **47**, 467.

YOSHIDA, H. and FUJISAWA, H. (1962) *Biochim. biophys. Acta*, **60**, 443.

ZERAHN, K. (1956) *Acta physiol. scand.*, **36**, 300.

XV

Formation and Drainage of the Cerebrospinal Fluid

H. DAVSON

THE main outlines of the formation and drainage of the cerebrospinal fluid were sketched and submitted to experimental test by Weed in his classical papers (1914–1935). According to these, the fluid is produced continuously within the ventricles by the choroid plexuses and, having passed out of the IVth ventricle into the large subarachnoid space called the cisterna magna, it flows over the surface of the brain to be drained away into the blood in the large dural sinuses (Fig. 1). This general concept thus raises two important physiological problems, namely the mechanism of production of the fluid by the choroid plexuses and the mechanism of drainage into the blood stream. We may discuss the second point first.

THE DRAINAGE MECHANISM

The site of absorption of the cerebrospinal fluid was recognized by Weed to be the microscopic arachnoid villi that projected through the dural and endothelial linings of the dural sinuses. According to Weed, these villi could be regarded as growths of arachnoid tissue into the blood vessel, a separation between blood, on the one hand, and cerebrospinal fluid on the other, being maintained by the presence of a continuous mesothelial covering on the villus. On this basis, drainage would be favoured by two physical factors, namely a difference of *hydrostatic pressure* forcing cerebrospinal fluid across the mesothelial lining of the villus, and a difference of *osmotic pressure* between the cerebrospinal fluid and the blood plasma, a difference that could be described as the colloid osmotic pressure resulting from the low concentration of proteins in the fluid by contrast with that in the blood plasma. A variety of measurements on the pressures within the cerebrospinal system—the fluid pressure—and the blood pressure in the dural sinuses concur in attributing a pressure differential favouring flow from the cerebrospinal fluid into the sinus (see, for example, Weed, 1935*b*; Bedford, 1935; Shulman, Yarnell and Ranshohoff, 1964). The second factor, namely the difference of colloid osmotic pressure, must be ruled out on both theoretical (Davson, 1956) and experimental (Welch and Friedman, 1960) grounds. Thus the mesothelial cap of the arachnoid villus was considered to be a membrane, permeable to the crystalloids of the cerebrospinal fluid and blood, but impermeable to the plasma proteins; as such, it would undoubtedly develop a difference of osmotic pressure favouring passage of cerebrospinal fluid into the plasma; in fact, the situation would be similar to that envisaged by Starling (1896) when describing the reabsorption of extracellular fluid into a venous capillary. The theoretical objection to the mechanism is that it makes no provision for the absorption of proteins from the cerebrospinal fluid; the fluid as secreted into the ventricles contains a definite concentration,

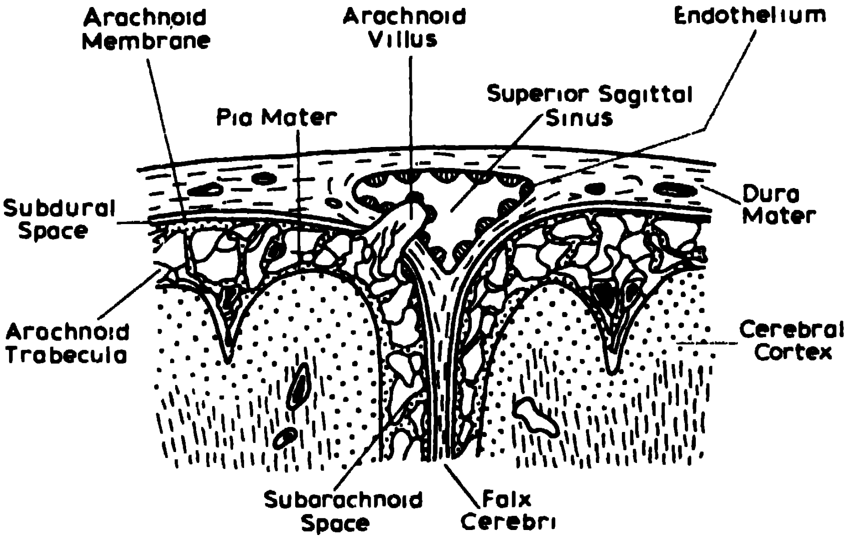


FIG. 1. Illustrating the meninges and an arachnoid villus projecting into the superior sagittal sinus. (After Weed, *Amer. J. Anat.*)

albeit small, of the individual plasma proteins (a total of about 17 mg/100 ml, according to a recent study by Hill, McKenzie, McGuckin, Goldstein and Svien, 1958). If the membrane covering the villus is to be impermeable to proteins, or to exert sufficient restraint on the passage of proteins to cause a difference of osmotic pressure, then such a restraint will lead to a filtering back of the protein in the cerebrospinal fluid, which will thus become more and more concentrated as absorption of the fluid proceeds, and ultimately the concentration of proteins in the fluid will rise to that in the plasma. Thus, this 'colloid osmotic absorption' is a feasible mechanism for absorption of extracellular fluid in, say, skeletal muscle, because there is a lymphatic system that carries away the proteins left behind during the absorptive process across the capillary wall. In the cerebrospinal system there are no lymphatics capable of doing this, so that the small quantities of protein that appear normally in the fluid, and the larger quantities that may appear pathologically, can only be returned to the blood by a bulk flow mechanism that does not exert an appreciable restraint on the protein constituents of the cerebrospinal fluid. The only anatomically feasible place for this absorption is, in my view, the arachnoid villus, provided it has sufficiently large pores on its surface to permit the ready passage of proteins, and even bigger particles such as red blood cells, since the rapid clearance of blood from the cerebrospinal fluid suggests that even these large bodies can escape without serious impediment (see, for example, Russell, 1932; Simmonds, 1953). The anatomical and physiological studies of Welch and Friedman and Welch and Pollay have resolved this difficulty; according to their histological studies, the villus can be described as a relatively coarse meshwork with a valvular function, such that when the cerebrospinal pressure rises above that in the blood vessel by a critical amount, the meshes open up to permit the flow of fluid; when the pressure-gradient falls below the critical value, the meshwork closes up and thus prevents a reflux of blood into the subarachnoid space. The process is illustrated schematically in Fig. 2. By excising pieces of dura containing numerous villi, Welch and Friedman (1960) were able to set up an artificial system with the dura separating two chambers. Under these conditions a definite 'opening pressure', of about 60 mm fluid, was required to induce flow from the cerebrospinal fluid side to the blood side; this depended on the surface tension, since addition of a detergent reduced the 'opening pressure' to zero. The rate of drainage was quite independent of the colloid osmotic pressure of the fluid on the blood side, as predicted; when particulate suspensions were put on the cerebrospinal fluid side these passed throughout without appreciable filtering back, even when monkey erythrocytes of diameter about 7 μ were the particles; only with polystyrene spheres of diameter 6·4 to 12·8 μ range was there exclusion, and the particles were found in the meshes of the villi (Welch and Pollay, 1961).

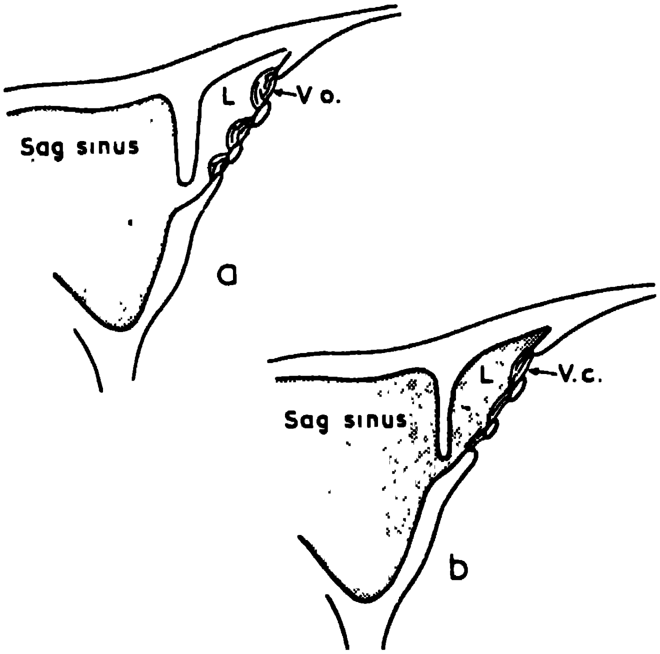


FIG. 2. Diagrammatic illustration of the situation and function of the arachnoid villi. When the meningeal pressure is sufficiently greater than that in the sinus, the villus meshwork is open as in (a). When the meningeal pressure is below that in the sinus, reflux of blood is prevented by collapse of the villus (b). (Welch and Friedman, *Brain.*)

SPINAL DRAINAGE

The problem as to whether there is drainage from the spinal cord is still an open one. In man and other mammalian species it is considered that the spinal canal, i.e. the extension of the

IVth ventricle into the cord, is little more than a potential space ending blindly in the filum terminale, so that flow of fluid out of the IVth ventricle is entirely by way of the foramina of Magendie and Luschka into the cisterna magna, and thence into the remaining subarachnoid spaces. In order that there should be a consistent flow of fluid in the spinal subarachnoid space there must be some drainage of fluid into the blood vessels of the cord or of the spinal roots, or perhaps a flow along the sheaths of the spinal roots. Such physiological evidence as there is, bearing on this point, militates against any considerable flow in the spinal subarachnoid space, the renewal of the fluid depending on chance variations in pressure in the different parts of the cerebrospinal system (Sachs, Wilkins and Sams, 1930; Grundy, 1963). Nevertheless, there is some anatomical evidence suggesting the presence of arachnoid villi in the spinal subarachnoid space, as illustrated by Fig. 3, from Welch and Pollay (1963). As the figure shows, in the region of the emerging root the arachnoid is reflected from the dura on to the intra-

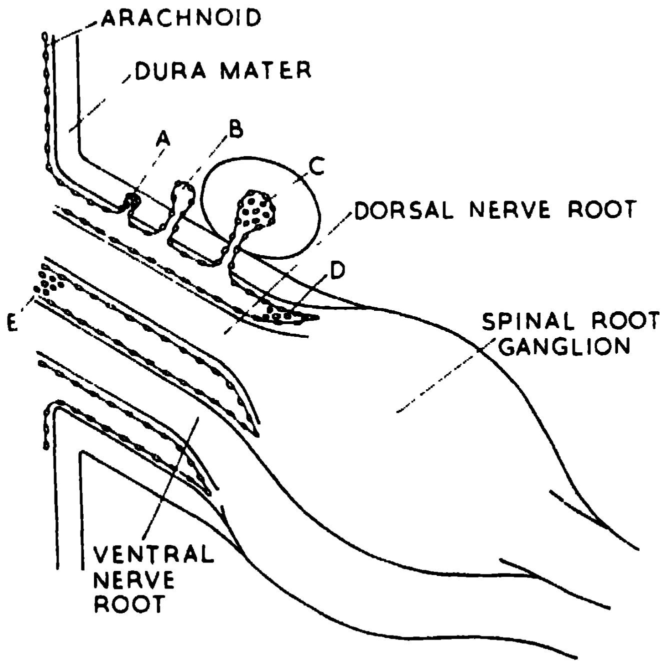


FIG. 3. Diagrammatic representation of the kinds of meningeal and vascular relationships found in spinal nerve roots. At A, growth of arachnoidal cells within the dura mater is represented. B represents complete penetration of dura mater by leptomeninx and C protrusion of arachnoidal tissue into a vein. At D and E arachnoidal proliferations are shown. (Welch and Pollay, *Anat. Rec.*)

dural surface of the roots to produce blind pockets. In these pockets, proliferations of arachnoidal tissue of various kinds (A to E in Fig. 3) are observed, characterized by the degree of invasion of the dura. Type C represents the equivalent of a cranial arachnoid villus, since it has penetrated an extradural vein. A similar spinal arachnoid villus was described in 1923 by Elman. As Welch and Pollay point out, it is quite likely that some of these proliferations may represent the pathways for escape of particulate matter so frequently observed in experimental studies designed to show pathways of emergence of fluid (cf. Brierley and Field, 1948; Woollam and Millen, 1958; Howarth, 1952; Howarth and Cooper, 1955).

The Spinal Canal

In one species, namely the rabbit, a flow of cerebrospinal fluid down the spinal canal has been demonstrated (Bradbury, Davson and Lathem, 1964). Thus, if, in this species, a dye is injected into the lateral ventricle, it appears rapidly at all places along the canal and later is found in the spinal subarachnoid space; the distribution is shown in Fig. 4, *top*, and the circumstance that the caudal space is strongly coloured whilst the intervening dorsal and lumbar regions are not, is strong evidence for an escape from the canal directly into the sacral subarachnoid space. (The colour in the cervical region is due

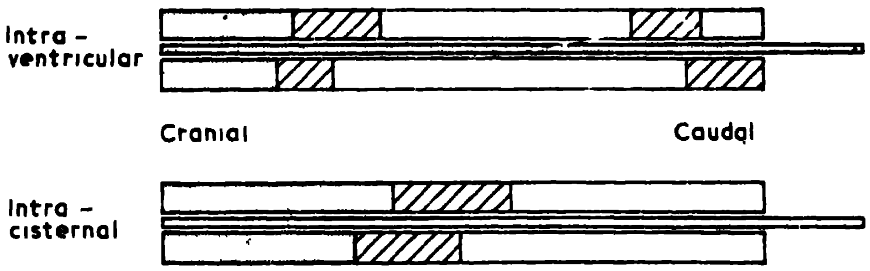


FIG. 4. Final distribution of dye following intraventricular injection and intracisternal injection. For each type of experiment, top horizontal column refers to dorsal subarachnoid distribution, middle column to distribution in central canal and filum terminale, and bottom column to ventral subarachnoid distribution. Stippled segments indicate marked coloration and cross hatched, some coloration visible. Distances as fractions of total spinal cord length. (Bradbury, Davson and Lathem, *J. Physiol.*)

to diffusion down the space from the cisterna magna.) The escape of dye appeared to take place through the filum terminale, since a transverse section of the caudal end of the cord always showed uncoloured nervous tissue between the central canal and the pia mater. The amount of flow is small (0·05 μl/min compared with an estimated total production of 10 μl/min), and is presumably brought about by the cilia that line the canal.

SECRETION OF THE FLUID

The choroid plexuses are essentially pouchings of the pial vascular system into the ventricles; as such, they consist of a highly vascular core covered with ependyma, which is now called the *choroidal epithelium.* Such a close juxtaposition of blood capillaries and epithelium is probably an ideal arrangement for the production of a secretion. The capillaries would produce a filtrate from plasma into the adjacent connective tissue spaces of the plexus, and this would be largely prevented from passing into the ventricle by the tightly packed epithelial cells which would act as transporting agencies, actively removing such constituents as sodium, chlorine, potassium, etc. from the filtrate and discharging them into the ventricular cavity, producing thereby local differences of osmotic pressure that would cause the passage of water. If this latter process were rapid enough, the osmotic pressure of the resulting secretion might be very close to that of the parent blood plasma. The most recent study on the relative osmolarities of plasma and cerebrospinal fluid indicate a small degree of hyperosmolarity of the cerebrospinal fluid (Bito and Davson, unpublished).

The demonstration that the choroid plexuses are the source of the cerebrospinal fluid we owe to Dandy and Blackfan (1914) and Frazier and Peet (1914), who produced an experimental hydrocephalus by plugging the aqueduct of Sylvius. If the choroid plexus of one lateral ventricle was removed, and both foramina of Monro (which lead from the lateral ventricles into the IIIrd ventricle) were blocked, one ventricle dilated and the other collapsed. At various times the role of the choroid plexuses as the sole source of the cerebrospinal fluid has been disputed (Hassin, 1947; 1948; Bering and Sato, 1963), and it has been suggested that the fluid is formed in a non-specific fashion in the tissues adjacent to the ventricles and subarachnoid space; but although some indirect evidence has been adduced in support of the claim that at least a part of the fluid is formed in this manner (Bering and Sato, 1963), the precise sites of formation and the mechanism of transport through the tightly packed tissue have not been adequately demonstrated or, for that matter, adequately discussed.

Proof that the choroid plexuses most probably give rise to the cerebrospinal fluid has been given by the study of the blood leaving a plexus (Welch, 1963), and by the chemical analysis of fluid collected directly from the surface of the exposed plexus (Ames, Sakanoue and Endo, 1964). Welch cannulated the large vein that drains approximately 75 per cent of the blood from the choroid plexus of the lateral ventricle of the rabbit. If the blood flowing through the plexus is the source of the secreted cerebrospinal fluid, then the venous blood collected should contain less water per unit volume than the arterial blood, i.e. the haematocrit value of the venous blood, expressed as the proportion of cells to whole blood, should be greater than the haematocrit value for arterial blood. Direct comparison revealed that some haemoconcentration did, indeed, occur, so that the ratio of the haematocrits was of the order of 1:15. If the rate of blood-flow through the plexus could be estimated, we would be able to deduce the rate of secretion of cerebrospinal fluid. By measuring cinematographically the rate of passage of oil particles injected into the blood, Welch deduced an arterial blood flow of some 2·8 µl/min per mg of plexus, and from this he computed a secretion rate of 0·37 µl/min per mg plexus, or 2·6 µl/min for one choroid plexus. This, in the rabbit, is probably about one third of the total production, so that a figure of about 8 µl/min could be taken as an indication of the rate of production of fluid.

By perfusing an artificial cerebrospinal fluid at a known rate into the ventricular system, draining the perfused fluid at the cisterna magna, the amount of fluid secreted in all the ventricles can be estimated accurately by incorporating inulin into the perfusion fluid and measuring the change in concentration that it undergoes on passage through the system (Pappenheimer, Heisey, Jordan and Downer, 1962). Thus, if C_{In} is the concentration in the fluid as it enters, and C_{Out} the concentration in the emerging fluid, the rate of secretion is given by:

$$\frac{C_{In} - C_{Out}}{C_{Out}} \times F_v$$

where F_v is the rate of perfusion. This is true approximately because inulin does not diffuse appreciably out of the fluid into the adjacent tissues, so that any change in its concentration is due to dilution by freshly secreted fluid, which, of course, contains no inulin. Applying this technique to the rabbit, Pollay and Davson (1963) obtained a figure of 8·8 µl/min, whilst more recently Bradbury and Davson (1964) have given a value of 10·1 µl/min. The fact that these estimates agree reasonably well with those of Welch is strong evidence that the choroid plexuses are a source of cerebrospinal fluid and may well be the *only*; but, of course, more accurate estimates of blood flow through the plexus would be required to justify a closer comparison. (In larger animals, as we should expect, the rates of production of fluid are higher; thus, in the cat it is 20 µl/min (Davson, Kleeman and Levin, 1962), in the dog, 50 µl/min (Oppelt, Patlak and Rall, 1964), and in the goat, 160 µl/min (Heisey, Held and Pappenheimer, 1962); in general, these rates correspond to about the same turnover rate, namely some 0·3 to 0·5 per cent per min.)

The second piece of evidence bearing on secretion of cerebrospinal fluid is derived from microanalysis of the freshly secreted fluid collected from the surface of the exposed plexus; the concentrations of sodium, chlorine, potassium, calcium and magnesium in this fluid are sufficiently similar to that of the fluid collected in bulk from the cisterna magna to justify the belief that the fluid so collected is, indeed, the primary cerebrospinal fluid (Ames *et al.*, 1964).

CHEMISTRY OF THE CEREBROSPINAL FLUID

We have described the process of formation of the fluid by the choroid plexuses as one of secretion, as opposed to a process of plasma filtration. This implies, essentially, that the epithelial cells, in transporting fluid and solutes from the blood to the fluid side of the plexus, are performing secretory work requiring a suitable supply of metabolic energy. This conclusion is reached by studying the chemistry of the cerebrospinal fluid; superficially it is similar to plasma except for the low concentration of proteins, and it is not surprising that it was taken for a simple filtrate of plasma at one time (Fremont-Smith, Thomas, Dailey and Carroll, 1931). However, modern chemical studies show that the *differences* from a plasma filtrate are significant; thus the concentrations of a few solutes in cerebrospinal fluid, plasma and a filtrate of plasma are shown in Table 1; these differences between the composition of the fluid and a plasma filtrate are significant analytically, and this is especially true of potassium, whose concentration is only 2·9 meq/litre compared with 4·35 meq/litre in a plasma filtrate; chlorine, with a concentration of 130 vs 111 meq/litre in a filtrate; urea (30 mg/100 ml compared with 50 mg/100 ml) and glucose (97 compared with 150 mg/100 ml).

TABLE 1. Approximate concentrations of certain ions and non-electrolytes in the cerebrospinal fluid of the rabbit compared with the concentrations in a filtrate from plasma

Substance	Cerebrospinal fluid	Plasma filtrate
Na	148 mequiv/litre	140 mequiv/litre
Cl	130 ,,	111 ,,
HCO_3	24 ,,	27 ,,
K	2·96 ,,	4·35 ,,
Urea	39 mg/100 ml	50 mg/100 ml
Glucose	97 ,,	150 ,,

CONSTANCY OF THE COMPOSITION

The cerebrospinal fluid has, as we see, a significantly different composition from that of blood plasma or its filtrate; and this is because active transport mechanisms operate across the choroidal epithelium, accelerating certain molecules and ions and retarding others. We may presume that the result of these processes is to produce a fluid whose composition is well adapted to act as an environment for the central neurones, or at any rate for those that come into close relationship with the fluid. It would not be surprising, then, if the composition of the cerebrospinal fluid were, to a large extent, independent of that of the plasma, unlike a plasma filtrate whose composition would vary directly as that of the plasma varied. To some extent this independence is realized. Thus, if we examine the concentration of potassium in the cerebrospinal fluid and plasma of a large number of human subjects (Bradbury, Stubbs, Hughes and Parker, 1963), we find that, although the concentration in the blood plasma varies between 3·5 and 6 milliequivalents per litre, the concentration in the cerebrospinal fluid varies over a much smaller range above and below 2·9 mequiv/litre, and shows no correlation at all with the plasma level (Fig. 5). In experimental animals this constancy can be demonstrated even more emphatically; thus one may maintain a high level of potassium in the blood of a rabbit for 24 hours and yet find the concentration in the cerebrospinal fluid unchanged (Bekaert and Demeester, 1951; Dunker, 1957; Bradbury and Davson, 1965). Essentially similar results are found with calcium and magnesium (Schain, 1964; Bradbury, 1965), and we may interpret this constancy on the basis of a transport from the blood by some sort of 'carrier mechanism' that is saturated at normal, or lower than normal, levels of the respective ions in the plasma.

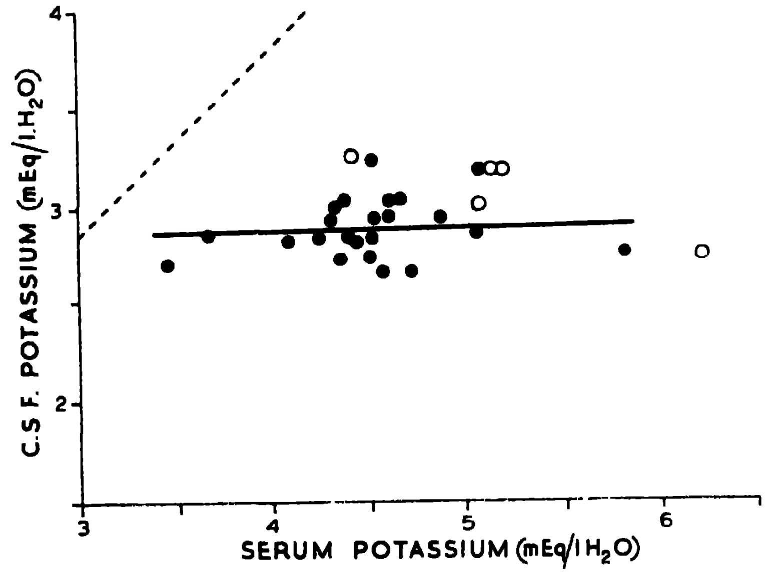


FIG. 5. The relation of lumbar C.S.F. potassium to serum potassium for subjects with normal barriers (●) and subjects with active tuberculous meningitis (○). The continuous line (——) is the regression line for those with normal barriers; the broken line (– – – –) follows the ratio for a dialysate of plasma. (Bradbury, Stubbs, Hughes and Parker, *Clin. Sci.*)

Blood-Cerebrospinal Fluid Barrier

This relative constancy of composition does not mean, however, that the cerebrospinal fluid is immune to changes in composition of the blood; far from it. Thus, increases in the total osmolarity of the blood, produced by intravenous injections of sodium chloride or mannitol, cause an abrupt withdrawal of water from the cerebrospinal fluid, lowering the intracranial pressure and raising the osmolarity of the fluid. Again, many foreign substances, when injected into the blood, find their way into the cerebrospinal fluid; some easily, and some with difficulty, so that we may speak of a *blood-cerebrospinal fluid barrier*, in the sense that there is an impediment to the passage from blood to fluid, an impediment that varies in degree from substance to substance. Some examples are shown in Fig. 6 where substances with different lipid-solubility have been examined; in these experiments a constant concentration of the substance was established in the plasma, and the concentration in the cerebrospinal fluid was determined at different intervals. The kinetics of penetration are remarkably regular and usually follow a simple exponential course, so that transfer-constants, indicating the rates of coming to equilibrium, can be computed. In Fig. 7

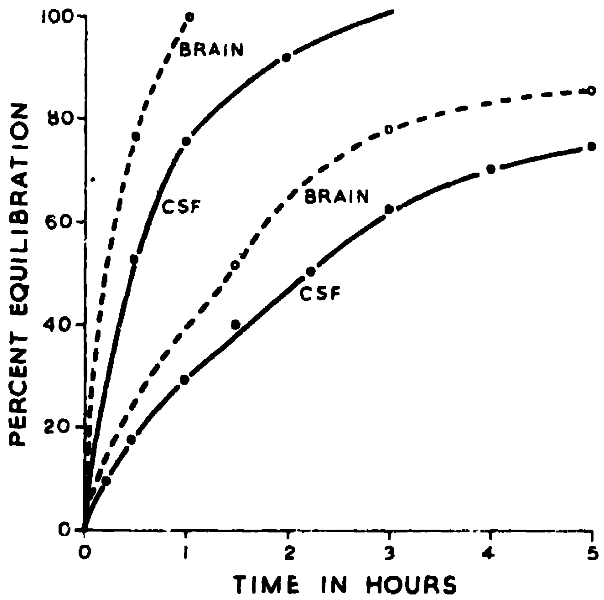


Fig. 6. Penetration of ethyl thiourea (*upper curves*) and thiourea (*lower curves*) into cerebrospinal fluid and brain. (Davson, *Ergebnisse der Physiologie.*)

the results of more recent studies on drugs have been plotted logarithmically; once again the importance of lipid-solubility becomes manifest.

It will be seen from Fig. 6 that the rate of equilibration of ethyl alcohol is very high, and this raises the question as to how this is brought about. If we assume that the alcohol in the blood equilibrates very rapidly with the water in the cells of the secreting choroidal epithelium, then the new fluid, as it enters the ventricles, will contain about the same concentration of alcohol as that maintained in the blood. As time passes, the alcohol-containing fluid will mix with, and ultimately replace, the original fluid, so that finally the concentrations in plasma and cerebrospinal fluid will become equal. The time-course of this process will clearly be determined by the rate of renewal

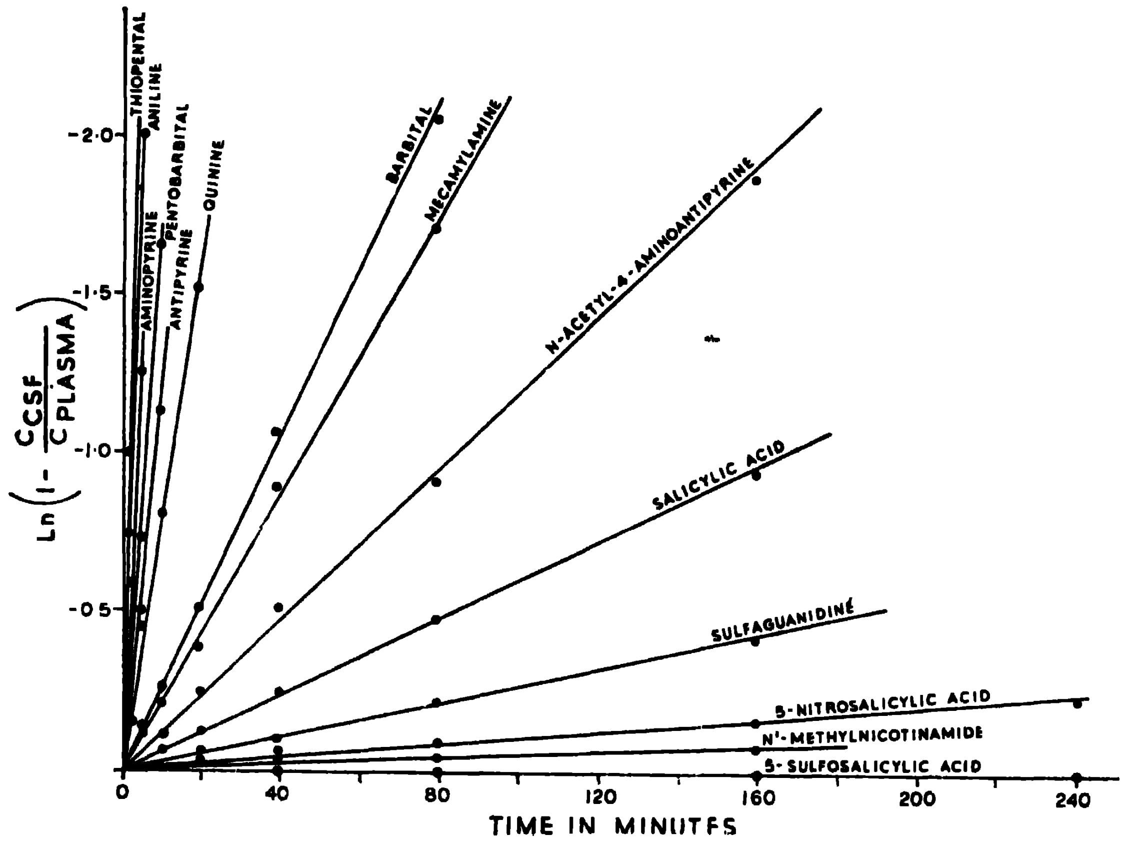


FIG. 7. Passage of drugs from the blood-stream into the cerebrospinal fluid (CSF) of dogs. C_{CSF} refers to the concentration of drug in CSF, and C_{PLASMA} to that in plasma after correction for the degree of protein binding. The relative rates of penetration into CSF are given by the slopes of the lines. (Schanker, *Advances in Drug Research.* Academic Press.)

of the cerebrospinal fluid, and this is a relatively slow process, the 'half-life' for renewal being approximately 2·5 hr, whereas Fig. 6 shows that the half-life for equilibration with blood is only a few minutes. Clearly, there must be an additional means of penetration, and this could be given by simple diffusion from the blood in the nervous tissue and thence into the cerebrospinal fluid. This will be more evident if we examine Fig. 8, where the relations of cerebrospinal fluid to blood and nervous tissue have been indicated schematically. From this it will be seen how

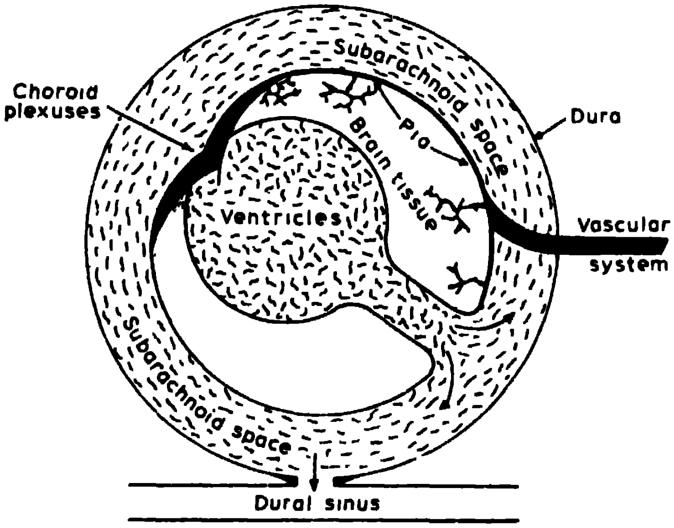


FIG. 8. Illustrating the relationships between cerebrospinal fluid, blood and nervous tissue. (Davson, *Ergebnisse der Physiologie.*)

closely the cerebrospinal fluid is related to the adjacent nervous and glial tissue, and since this has its own blood supply in higher vertebrates, the passage of material from the blood can, at least theoretically, take place in two distinct fashions; first, by entry into the secreted fluid, and secondly by escape from blood into the extracellular fluid of the brain and thence into the ventricles and subarachnoid spaces across the ependymal and pial linings respectively. In order that the second route from the brain can contribute, however, its concentration must be greater than that in the cerebrospinal fluid at any time during the equilibration process. With ethyl alcohol this becomes very manifest, since in the shortest time feasible for study (2 minutes) the brain was found to have completely equilibrated with the blood, whilst the cerebrospinal fluid was far behind. Fig. 9 shows the situation with two less rapidly penetrating substances, thiourea and ethyl thiourea. In both cases the brain was in advance of the cerebrospinal fluid.

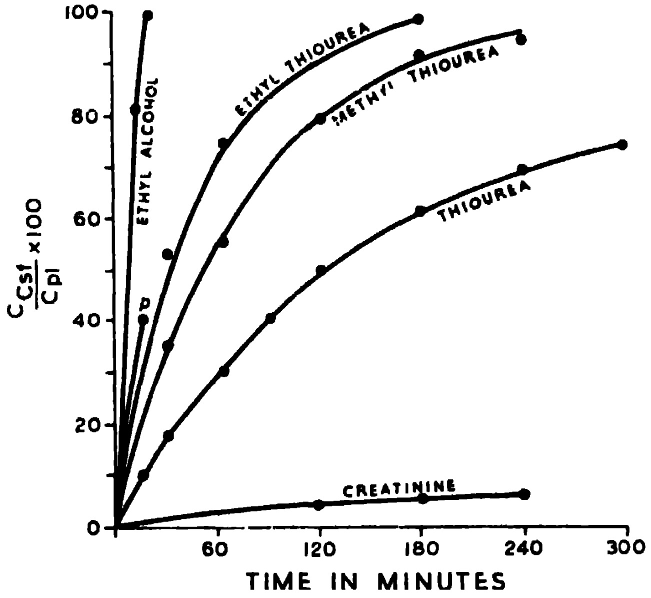


FIG. 9. Illustrating the blood-C.S.F. barrier. A constant concentration of the substances was maintained in the plasma and the concentration in the cerebrospinal fluid determined at different times. (Davson, *J. Physiol.*)

BLOOD-BRAIN BARRIER

The consideration raised above shows us that, if we are concerned with the relation of the cerebrospinal fluid to the blood, we must also concern ourselves with the relation of the brain to the blood, and the relation of the brain to the cerebrospinal fluid. The former relationship we may summarize by the concept of the ***blood-brain barrier***. Thus, when a substance is injected into the blood, the rapidity with which it enters the brain tissue may be great or small, depending on the special characteristics of the substance. As we have seen, lipid-soluble substances like ethyl alcohol penetrate very rapidly indeed; on the other hand, there are many substances that penetrate so slowly as to be virtually excluded, e.g. inulin, sucrose, iodide, thiocyanate, p-aminohippurate, penicillin. Other substances, like urea, creatinine, various ions such as ^{24}Na, ^{42}K, ^{45}Ca, occupy an intermediate position. It is this *selectivity* that has given rise to the modern concept of the blood-brain barrier; and it is in such marked contrast to what is observed in other tissues such as skeletal muscle, where such substances as sucrose and inulin equilibrate rapidly with the extracellular fluid of the tissue, that one is tempted to look for a special covering to the capillaries of the central nervous system that could impart to them this selectivity. There is some evidence that it is the astroglial end-feet that are responsible (see, for example, Clemente and Holst, 1954).

EXTRACELLULAR SPACE OF BRAIN

The blood-brain barrier is most marked for large, highly water-soluble, molecules, such as sucrose and inulin; these are typically 'extracellular substances' so that their passage from blood into brain is essentially a passage out of the capillaries into the extracellular space; thus, at complete equilibrium, the amount that was found in the tissue as a whole would depend on the volume of the extracellular space. For example, if this space constituted some 15 per cent of the tissue, as in skeletal muscle, we should expect, at equilibrium, to find the concentration in the tissue, expressed as millimoles per kg of tissue, to be only 15 per cent of the concentration in the plasma. If the volume of the extracellular fluid were only 1 per cent, then we should expect the concentration to be only 1 per cent of that in the plasma. Thus, even if there were no barrier, in the sense that inulin and sucrose passed very rapidly out of the blood stream into the tissue, there would appear to be a barrier because we should find so little sucrose or inulin in the brain by comparison with that in the plasma. Expressed in another way, the apparent barrier would be due to the circumstance that there was 'nowhere to go' for the sucrose or inulin. It has been seriously argued that this is, indeed, the reason for the blood-brain barrier, an argument that has been supported by the electron-microscopical appearance of brain which suggests that the extracellular space would be very small by comparison with other tissues (see, for example, Horstmann and Meves, 1959). This argument would be very cogent were there not a way of proving that there is, indeed, 'somewhere to go' for inulin, sucrose, p-aminohippurate, and other substances that exhibit the 'barrier phenomenon'. This may be done by soaking the tissue in artificial cerebrospinal fluid containing the solute (Davson and Spaziani, 1959), or by perfusing the cerebrospinal system (Davson, Kleeman and Levin, 1961; Rall, Oppelt and Patlak, 1962). Under these conditions the solute is not inhibited by the blood-brain barrier from penetrating the tissue, and it comes to equilibrium to give spaces of 10–15 per cent, spaces that may very well be the true extracellular space of the brain. The low values for extracellular space deduced from electron-microscopy are most probably due to post-mortem and fixation changes, since recently Van Harreveld and Crowell (1964) have developed methods for rapid freezing and substitution fixation; their application to the brain suggests that the extracellular spaces are, indeed, much larger than hitherto thought.

BRAIN-CEREBROSPINAL FLUID RELATIONS

We have indicated that substances like sucrose and inulin can be made to penetrate the brain by perfusing the ventricles, i.e. by allowing a non-vascular route. Systematic studies of the passage of substances from the cerebrospinal fluid to the brain (Pappenheimer, Heisey and Jordan, 1961; Davson, Kleeman and Levin, 1962; Heisey, Held and Pappenheimer, 1962; Bradbury and Davson, 1964) suggest that the ependymal and glial linings of the system do not offer a serious impediment to the passage of most solutes; certain substances do, indeed, escape much more rapidly than others, but these rapid rates of escape are predominantly characteristic of the blood-brain and blood-cerebrospinal fluid barriers. Thus ethyl alcohol escapes very rapidly indeed from the cerebrospinal fluid (Davson, Kleeman and Levin, 1962) and this is primarily because it escapes from the brain tissue into the blood, thereby maintaining a high gradient between cerebrospinal fluid and the tissue. Sucrose, by contrast, passes very slowly from the tissue into the blood, and therefore accumulates in the tissue and thus soon reduces the gradient of concentration between the cerebrospinal fluid and tissue to negligible proportions, bringing the escape to a virtual stop. Certain substances such as Diodrast, p-aminohippurate, iodide and thiocyanate escape very rapidly, but this is because of the activity of the choroid plexuses, which have a special mechanism for the removal of these substances from the cerebrospinal fluid (Pappenheimer, Heisey and Jordan, 1961; Becker, 1961; Welch, 1962; Pollay and Davson, 1963; Edström, 1964). The fact that exchanges of material may take place with relative ease across the ependyma is of great theoretical significance, which may be illustrated by the example of potassium. As we have seen, the concentration of this ion in the cerebrospinal fluid is only some 2·9 mequiv/litre, compared with some 4–5 mequiv/litre in plasma. The cerebrospinal fluid has quite a long 'life' in the ventriculo-subarachnoid compartments, so that if the extracellular fluid of brain has a different concentration of potassium from that in the cerebrospinal fluid, we may expect the concentration in the cerebrospinal fluid to change as the fluid passes from ventricles to subarachnoid spaces. Let us assume, for example that the concentration in the extracellular fluid is the same as that in a filtrate of plasma, as with other tissues; clearly, then, the concentration of potassium in the subarachnoid fluid should be higher than that in the ventricles; in fact, however, the fluid removed from the cortical subarachnoid space of the dog has a *lower* concentration than that in the ventricle or cisterna magna. Thus the mean values for plasma, cisterna magna and cortical subarachnoid fluid were 4·35, 2·98 and 2·58 mequiv/litre respectively (Bito and Davson, 1965). [This suggests that the extracellular fluid of brain has a lower concentration of potassium than in the cerebrospinal fluid, and thus a very much lower concentration than in a plasma filtrate; this means that the extracellular fluid of the brain must be recognized as a secretion with its composition maintained by active transport mechanisms.] It may well be that the astrocytes, whose end-feet cover the brain capillaries, are responsible for the elaboration of this extracellular fluid. So far as many physiologically important ions are concerned, e.g. potassium, calcium, magnesium, chlorine and so on, the concentrations are maintained at levels similar to, but not necessarily exactly equal to, those in the cerebrospinal fluid, so that the two fluids together constitute a medium of specific and relatively constant composition bathing the neurones of the central nervous system. Evidence supporting this concept is provided by perfusing an artificial cerebrospinal fluid through the ventricles of a rabbit, the fluid being driven into a lateral ventricle and withdrawn from the cisterna magna. In different experiments a different concentration of potassium was employed in the perfusion fluid, and the emerging fluid was analysed for this ion. The results are shown in Fig. 10, where the concentration of the emergent fluid is plotted against time; each curve corresponds to an experiment with a different concentration of potassium, indicated by the figures at the right (0, 2·9, 5, 7, 10 mequiv/litre). If we look at the plot for 2·9 mequiv/litre, i.e. for the normal concentration in the rabbit's cerebrospinal fluid, we see that it is a straight line, indicating that the concentration remains the same in inflowing and outflowing fluid. If we look at the curve for 10 mequiv/litre, we see that the concentration rises with time; in the first minutes the fluid emerging is pure normal cerebrospinal fluid, with concentration of 2·9 mequiv/litre; as time goes on, the high-potassium fluid comes through, but the concentration never

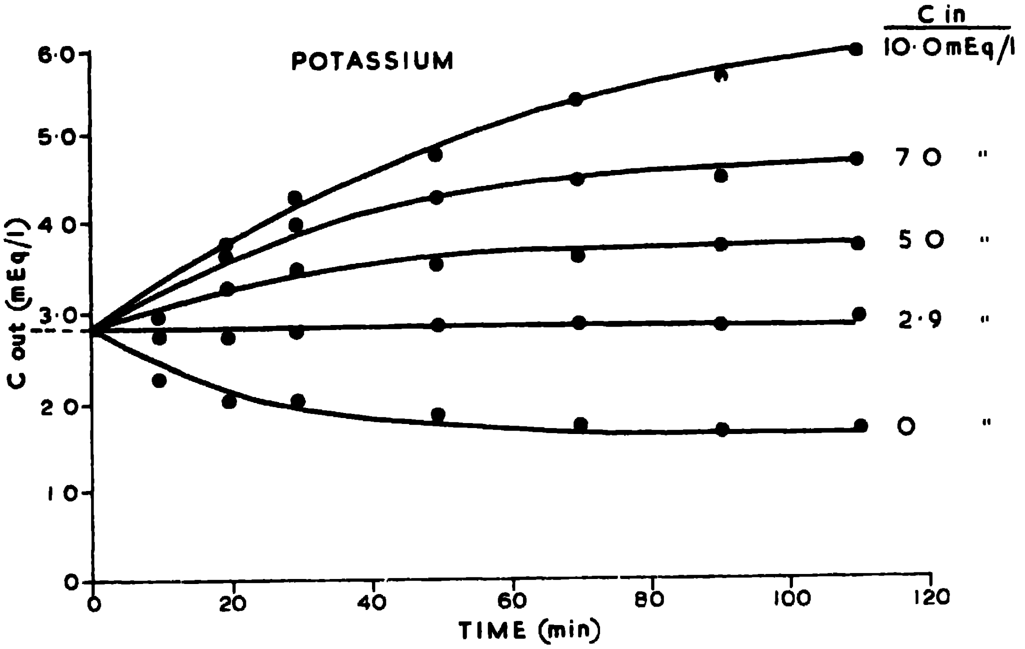


FIG. 10. Effect of ventriculo-cisternal perfusion with Ringer's solution containing different concentrations of K $^+$. Ordinate: Concentration of K $^+$ in outflowing fluid. Abscissa: Time from beginning of perfusion. The different concentrations of K $^+$ in the perfusion fluids (C_{in}) are indicated at the right of each curve. (Davson and Bradbury, *Progress in Brain Research*. Elsevier.)

reaches that in the inflowing fluid. This means that potassium diffuses into the adjacent tissues when the concentration of potassium is higher than the normal fluid value. If we look at the curve for zero potassium in the inflowing fluid, we see that, as before, the initial samples of the emergent fluid contain 2·9 mequiv/litre, but as the new fluid comes through, the concentration falls, but it never falls to zero, indicating that potassium now comes from the nervous tissue into the fluid. Thus, exchanges of potassium between cerebrospinal fluid and nervous tissue are demonstrated by these experiments to be possible; moreover, they occur whenever the concentration in the cerebrospinal fluid is different from 2·9 mequiv/litre, and this strongly suggests that the extracellular fluid of the tissue has this concentration too. Analysis of the kinetics of these exchanges shows that they could be quite well accounted for on the basis of passive movements across the ependyma, i.e. there is no need to assume the existence of active processes operating across this membrane.

It was mentioned above that the concentration of potassium in the cortical subarachnoid fluid was lower than that in the cisterna magna, presumably because the extracellular fluid also has a lower concentration, and it was imposing this on its adjacent cerebrospinal fluid. The situation with urea is interesting in this context; the concentration in the cisternal cerebrospinal fluid is lower than in plasma, whilst there is good evidence that the concentration in the brain extracellular fluid is about equal to that in plasma (Bradbury and Coxon, 1962; Kleeman, Davson and Levin, 1962); consequently we can expect the concentration in the cortical subarachnoid fluid to be higher than in the cisternal fluid; and this has been found (Bito and Davson, 1965). Thus the concentrations were 20·5, 18 and 20·2 mg/100 ml in plasma, cisternal and cortical subarachnoid fluids respectively. The behaviour of potassium and urea are therefore in marked contrast, and illustrate the difference between a substance whose concentration in cerebrospinal and extracellular fluids is rigidly controlled (potassium), and one whose concentrations are not so rigidly controlled, but apparently vary passively with the concentration in the blood.

To conclude, it would seem that the adequate control over the composition of the environment of the brain cells is only achieved, when this control is important, by control over both the cerebrospinal fluid and the extracellular fluid; in the one case the specialized ependymal cells of the choroid plexuses are responsible, and in the other it may be the astrocytes lining the capillaries that control the composition of the extracellular fluid. When control is apparently unimportant, as with urea, then it may happen that the extracellular fluid has a different concentration from that in the cerebrospinal fluid, and this leads to differences in concentration within the cerebrospinal fluid cavities, the 'older' cerebrospinal fluid in the cortical subarachnoid space having a concentration nearer that in the extracellular fluid, which itself has a concentration about equal to that in a plasma filtrate.

REFERENCES

AMES, A., SAKANOUE, M. and ENDO, S. (1964) *J. Neurophysiol.*, **27**, 672.
BECKER, B. (1961) *Amer. J. Physiol.*, **201**, 1147.
BEDFORD, T. H. B. (1935) *Brain*, **58**, 427.
BEKAERT, J. and DEMEESTER, G. (1951) *Arch. int. Physiol.*, **59**, 393.
BERING, E. and SATO (1963) *J. Neurosurg.*, **20**, 1050–63.
BITO, L. and DAVSON, H. (1965) *J. Physiol.*, **179**, 73–74 P.
BRADBURY, M. W. B. and COXON, R. V. (1962) *J. Physiol.*, **163**, 423.
BRADBURY, M. W. B., STUBBS, J., HUGHES, I. E. and PARKER, P. (1963) *Clin. Sci.*, **25**, 97.
BRADBURY, M. W. B., DAVSON, H. and LATHEM, W. (1964) *J. Physiol.*, **172**, 16 P.
BRADBURY, M. W. B. and DAVSON, H. (1964) *J. Physiol.*, **170**, 195.
BRADBURY, M. W. B. and DAVSON, H. (1965) *J. Physiol.* In preparation.
BRADBURY, M. W. B. (1965) *J. Physiol.*, **179**, 67 P.
BRIERLEY, J. B. and FIELD, E. J. (1948) *J. Anat.* London. **82**, 153.
CLEMENTE, C. D. and HOLST, E. A. (1954) *Arch. Neurol. Psychiat.* Chicago, **71**, 66.
DANDY, W. E. and BLACKFAN, K. D. (1914) *Amer. J. Dis. Child.*, **8**, 406.
DAVSON, H. (1956) *The Physiology of the Ocular and Cerebrospinal Fluids.* Churchill, London.
DAVSON, H. and SPAZIANI, E. (1959) *J. Physiol.*, **149**, 135.
DAVSON, H., KLEEMAN, C. R. and LEVIN, E. (1961) *J. Physiol.*, **159**, 67 P.
DAVSON, H., KLEEMAN, C. R. and LEVIN, E. (1962) *J. Physiol.*, **161**, 126.
DUNKER, E. (1957) *Pflüg. Arch. ges. Physiol.*, **265**, 66.
EDSTRÖM, R. (1964) *Acta physiol. scand.*, **61**, 212.
ELMAN, R. (1923) *Johns Hopk. Hosp. Bull.*, **34**, 99.

FRAZIER, C. H. and PEET, M. M. (1914) *Amer. J. Physiol.*, **35**, 268.

FREMONT-SMITH, F., THOMAS, G. W., DAILEY, M. E. and CARROLL, M. P. (1931) *Brain*, **54**, 303.

GRUNDY, H. F. (1963) *J. Physiol.*, **163**, 457.

HASSIN, G. B. (1947) *J. Neuropath.*, **6**, 172.

HASSIN, G. B. (1948) *J. Neuropath.*, **7**, 432.

HEISEY, S. R., HELD, D. and PAPPENHEIMER, J. R. (1962) *Amer. J. Physiol.*, **203**, 775.

HILL, N. C., MCKENZIE, B. F., MCGUCKIN, W. F., GOLDSTEIN, N. P. and SVIEN, H. J. (1958) *Proc. Mayo Clinic*, **33**, 686.

HORSTMANN, E. and MEVES, H. (1959) *Z. Zellforsch.*, **49**, 569.

HOWARTH, F. (1952) *Brit. J. Pharmacol.*, **7**, 573.

HOWARTH, F. and COOPER, E. R. A. (1955) *Acta Anat.*, **25**, 112.

KLEEMAN, C. R., DAVSON, H. and LEVIN, E. (1962) *Amer. J. Physiol.*, **203**, 739.

OPPELT, W. W., PATLAK, C. S. and RALL, D. P. (1964) *Amer. J. Physiol.*, **206**, 247.

PAPPENHEIMER, J. R., HEISEY, S. R. and JORDAN, E. F. (1961) *Amer. J. Physiol.*, **200**, 1.

PAPPENHEIMER, J. R., HEISEY, S. R., JORDAN, E. F. and DOWNER, J. De C. (1962) *Amer. J. Physicl.*, **203**, 763.

POLLAY, M. and DAVSON, H. (1963) *Brain*, **86**, 137.

RALL, D. P., OPPELT, W. W. and PATLAK, C. S. (1962) *Life Sciences*, No. 2, 43.

RUSSELL, W. R. (1932) *Brain*, **55**, 549.

SACHS, E., WILKINS, H. and SAMS, C. F. (1930) *Arch. Neurol. Psychiat.* Chicago, **23**, 130.

SCHAIN, R. J. (1964) *Arch. Neurol.*, **11**, 330.

SHULMAN, K., YARNELL, P. and RANSHOHOFF, J. (1964) *Arch. Neurol.*, **10**, 575.

SIMMONDS, W. J. (1953) *Austral. J. exp. Biol. med. Sci.*, **31**, 77.

STARLING, E. H. (1896) *J. Physiol.*, **19**, 312.

VAN HARREVELD, A. and CROWELL, J. (1964) *Anat. Rec.*, **149**, 381.

WEED, L. H. (1914) *J. med. Res.*, **31**, 51.

WEED, L. H. (1917*a*) *Anat. Rec.*, **12**, 461.

WEED, L. H. (1923) *Amer. J. Anat.*, **31**, 191.

WEED, L. H. (1935*a*) *Brain*, **58**, 383.

WEED, L. H. (1935*b*) *Amer. J. Physiol.*, **114**, 40.

WELCH, K. and FRIEDMAN, V. (1960) *Brain*, **83**, 454.

WELCH, K. and POLLAY, M. (1961) *Amer. J. Physiol.*, **201**, 651.

WELCH, K. (1962) *Amer. J. Physiol.*, **202**, 757.

WELCH, K. and POLLAY, M. (1963) *Anat. Rec.*, **145**, 43.

WELCH, K. (1963) *Amer. J. Physiol.*, **205**, 617.

WOOLLAM, D. H. M. and MILLEN, J. W. (1958) 'Observations on the production and circulation of the cerebrospinal fluid'. Ciba Symp., p. 124. Churchill, London.

XVI

Nervous Control of Limb Blood Flow

I. C. RODDIE

IN many ways it is surprising how much we know of the ways in which blood flow in human limbs may be regulated through vasomotor nerves. It makes a contrast with how little we know of the behaviour of the vasomotor nerves which supply the other parts of the human body. This difference can probably be explained by the excellence of human limbs as preparations for vascular experiments. They provide paired, fairly symmetrical vascular beds perfused with an identical pressure head of blood. Experimental changes can always be compared with those in the opposite control limb and the blood vessels and nerves are relatively accessible. Blood flow may be measured without restraining or anaesthetizing the subject; this is important when studying the normal behaviour of vasomotor reflexes (McArdle and Roddie, 1958). The chief disadvantage of using human limbs is that it precludes any gross interference with the tissues. However the discipline provided by these limitations ensures that the tissues are in a highly physiological condition.

The information that we have on the vasomotor control of limb blood vessels has been gained from fairly simple experiments and consists mainly of inferences drawn from the effects of certain procedures on limb blood flow. It is these experiments I want to review in this paper.

BLOOD VESSELS IN SKIN

Claude Bernard (1852) was probably the first physiologist to show that nerves could influence blood vessels. In simple and elegant experiments on the rabbit he found that if the cervical sympathetic nerves were divided on one side, the vessels in the corresponding ear dilated. On account of the greater flow of blood, the side of the section became warmer than the normal ear. Stimulating the head end of the divided sympathetic nerve caused vasoconstriction in the ear vessels and cooling of the ear. He concluded that the sympathetic nerves exerted a tonic vasoconstrictor action on the ear blood vessels. Using the same experimental approach it can be shown that the pinna of the human ear is similarly supplied with vasoconstrictor nerves (Fig. 1). Blocking the great auricular nerve with local anaesthetic causes flushing and increased warmth in the corresponding ear (Blair, Glover and Roddie, 1961; Fox, Goldsmith, Kidd and Lewis, 1962). The blood vessels of the skin of these extremities are thus held under a high degree of vasoconstrictor tone. It has not yet been possible to stimulate human vasomotor nerves through the skin. A similar pattern of vasomotor control is seen in the skin of the hands and feet. Adson and Brown (1925) at the Mayo Clinic found that following lumbar or cervical sympathectomy there was an increase in heat elimination from the hands and feet indicating that the sympathetic nerves were normally exerting a tonic vasoconstrictor influence on the blood vessels of the extremities.

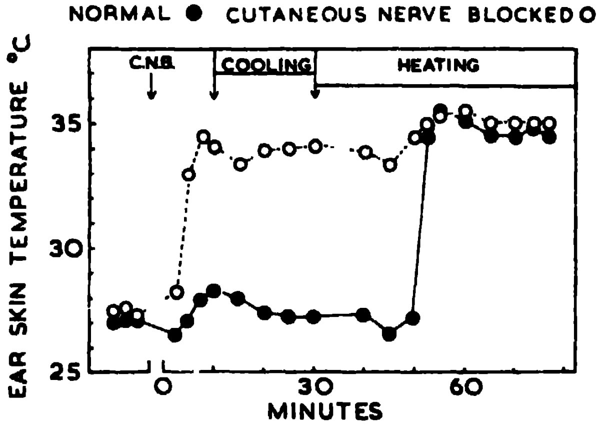


Fig. 1. Changes in the skin temperature of the normal and nerve blocked ear during body cooling and heating. At C.N.B., the great auricular nerve to one side was blocked as it appeared behind the posterior border of the sternomastoid muscle. (Blair, Glover and Roddie, 1961.)

A more quantitative estimate of the degree of vasoconstrictor tone in the hands and feet was made by Walker, Lynn and

Barcroft in 1950. They measured hand and foot blood flow by venous occlusion plethysmography before and for several days following a one-sided sympathectomy. For the first few days after the operation, the blood flow to the sympathectomized side was increased due to release of vasoconstrictor tone. Over the next few weeks the blood flow on the sympathectomized side gradually fell towards that of the normal side but the mechanism of the return of tone to sympathectomized vessels is not fully understood.

The degree of vasoconstrictor tone in the extremities is chiefly varied to meet the needs of temperature regulation, an increase in body temperature causing vasodilatation by release of vasoconstrictor tone and a fall in body temperature causing vasoconstriction by increasing vasoconstrictor tone.

The pattern of vasomotor innervation in the skin of the extremities of the limbs is different from that seen in the skin of the more proximal parts. This was first noticed by Grant and Holling in 1938. They found that blocking the cutaneous nerves which carry the sympathetic fibres to the skin of the forearm did not cause any flushing or warming of the forearm skin. However, it did prevent the vasodilatation and sweating which normally occurred in the forearm when the body was heated. The experiment shown in Fig. 2 illustrates the pattern of vasomotor control of the blood vessels in forearm skin. Blood flow was measured in both forearms by venous occlusion plethysmography, the subject sitting comfortably in a warm room. At the arrow marked C.N.B., the cutaneous nerves which carry the sympathetic fibres to forearm skin were localized by faradic stimulation and blocked with local anaesthetic. As found by Grant and Holling (1938) and Edholm, Fox and Macpherson (1957) this caused no change in blood flow on either side. It can be concluded that the vessels in forearm skin unlike those in the extremities are not subjected to vasoconstrictor tone when the subject is comfortably warm. When the subject was cooled by immersing his feet in cool water and directing a cold blast of air over his body, the blood flow fell on the normal side but not on the nerve blocked side. It is clear that vasomotor fibres were responsible for the fall in flow. It seemed that the nerves concerned were vasoconstrictor fibres; treating the forearm with bretylium tosylate (Blair, Glover, Kidd and Roddie, 1960), which blocks vasoconstrictor fibres, prevented the fall. The degree of fall is of the same order as that produced by Edholm, Fox and Macpherson (1957) when they completely suppressed the cutaneous circulation in the forearm by iontophoresis of adrenaline into the skin. This suggests that the vasoconstrictor fibres can completely suppress the skin circulation. However, when the subject is comfortably warm there is no tonic vasoconstrictor influence on the vessels in forearm skin.

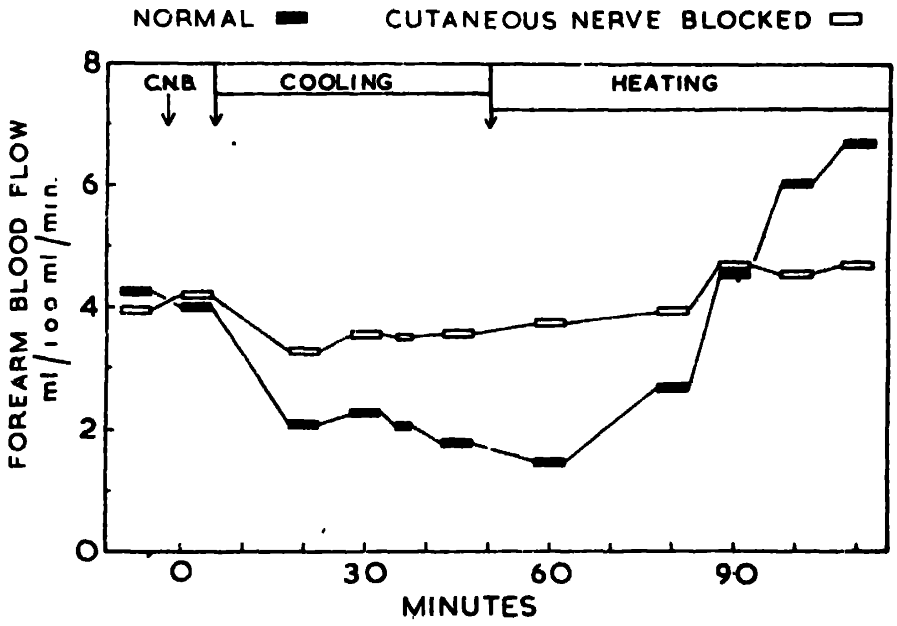


FIG. 2. The effects of body cooling and heating on blood flow in the normal and the cutaneous nerve blocked forearm. C.N.B.; cutaneous nerves on left side blocked with local anaesthetic. (Blair, Glover and Roddie, 1960.)

When the subject was subsequently heated by immersing his feet and calves in warm water and wrapping him in blankets the blood flow on the normal side rose, after a certain latency, to a level considerably in excess of that on the side in which the cutaneous nerves had been blocked. Vasomotor fibres running with the cutaneous nerves were clearly responsible for the vasodilatation on the normal side. The vasodilatation could not be explained by release of vasoconstrictor tone. Full release of vasoconstrictor tone would only increase the flow level to that seen in the nerve blocked forearm. It must therefore have been mediated by nerves whose activity results in vasodilatation, i.e. vasodilator nerves.

Vasodilator nerves act by releasing acetylcholine at their nerve endings. If the increase in flow with heating were due to the action of vasodilator fibres, treating the forearm with atropine should block the increase. This was found to be true to some extent (Roddie, Shepherd and Whelan, 1957). Fig. 3 shows plethysmograms from an experiment where forearm flow was measured on both sides. At the beginning of the experiment blood flow was nearly equal on the two sides as judged by the slope of the plethysmographic curves. When acetylcholine was infused into the left brachial artery in a strength of 20 μg/min (second panel) the blood flow to the left forearm became much greater as shown by increased steepness of the inflow curve. Ten minutes later, atropine was infused into the artery, 0·3 mgm being given over 3 minutes. Atropine by itself had no effect on the level of forearm blood flow (panel 3). However, it completely blocked the effect of a subsequent dose of acetylcholine. As can be seen from panel 4, the infusion of acetylcholine into the left brachial artery had no effect. The subject was then heated by immersing his feet and legs in warm water. After about 20 minutes of such heating, the blood flow on the normal (right) side has risen in the usual manner as can be seen from the increased steepness of the sample plethysmogram. However

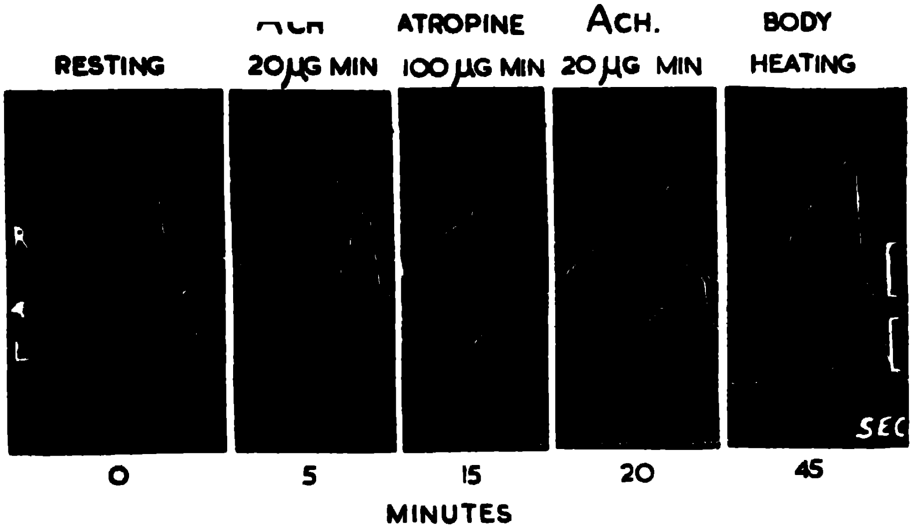


FIG. 3. Plethysmographic records of forearm blood flow before and during intra-brachial infusions of acetylcholine and atropine and during subsequent body heating. *Upper trace*: control forearm; *lower trace*: injected forearm. Time marker, seconds. (Roddie, Shepherd and Whelan, 1957.)

the blood flow on the atropinized forearm did not increase in a similar fashion. Atropine delayed and reduced the vasodilatation. This showed that nerves which were responsible for the vasodilatation in the forearm skin during body heating were cholinergic.

The close association of the onset of sweating with that of the vasodilatation in the forearm during body heating suggested that these two phenomena might be linked. Fox and Hilton (1958) have found evidence to support this. They identified bradykinin-forming enzyme activity in sweat collected from the hand and forearm. They found that the amount of bradykinin in the subcutaneous tissue space in the forearm increased when the body was heated and suggested the following hypothesis. Excitation of the sweat glands by cholinergic nerves results in the formation of bradykinin-forming enzyme by the glands. Some of this enzyme escapes from the glands into the periglandular tissue space where it breaks down tissue space protein into bradykinin, a vasodilator polypeptide. Bradykinin thus contributes to the cutaneous vasodilatation. If this were the case, the nervously mediated vasodilatation in response to heating would be a consequence of the nervous excitation of the sweat glands rather than to a direct action of the sympathetic fibres on the vessels.

Be that as it may, the increase in blood flow in the skin of the forearm during body heating unlike that in the skin of the extremities is due to an active vasodilator mechanism and not to release of vasoconstrictor tone.

Blood Vessels in Muscle

(1) *Vasoconstrictor fibres*

The pattern and function of the vasomotor innervation of blood vessels in muscle seem to be quite different from that in skin. Vasoconstrictor fibres to the blood vessels in human muscle were first described by Barcroft, Bonnar, Edholm and Effron in 1943. The experiments they carried out were as follows. Blocking the radial ulnar and median nerves to the forearm caused a 2–3 fold increase in forearm blood flow as measured by venous occlusion plethysmography. The increase in flow did not seem to be in skin since suppression of the cutaneous circulation by the iontophoresis of adrenaline into the skin did not prevent it and blocking the cutaneous nerves only did not produce it. It was not due to paralysis of the skeletal muscle, since blocking the radial, ulnar and median nerves in a patient whose arm had previously been sympathectomized did not cause an increase in flow. It was concluded that the blood vessels in skeletal muscle are normally subjected to considerable vasoconstrictor tone by sympathetic fibres which are distributed to the muscle in the motor nerves.

The vasoconstrictor fibres to muscle blood vessels are not concerned in temperature regulatory reflexes; temperature regulatory reflexes seem to be confined to the skin (Barcroft, Bock, Hensel and Kitchin, 1955; Edholm, Fox and Macpherson, 1956; Roddie, Shepherd and Whelan, 1956). There is evidence that the activity of these fibres is increased in muscular exercise in the muscles not actually involved in the exercise (Blair, Glover and Roddie, 1961). In the exercising muscles this effect is masked by the powerful local effect of metabolites. Fig. 4 shows plethysmograms of forearm blood flow, the subject lying supine on a platform and his legs placed on the pedals of a bicycle ergometer. At the arrow the subject began to exercise his legs on the ergometer and after about one minute the forearm flow fell as shown by the slopes of the plethysmograms becoming less steep. Fig. 5 shows the role of vasomotor nerves in this response. Blood flow was simultaneously measured in both forearms. The side represented by the open circles was normal but the other side whose blood flow is represented by the closed circles had the radial, ulnar and median nerves blocked with local anaesthetic about the elbow. The higher blood flow on this side was due to release of vasoconstrictor tone. In the 5 minute period between the two vertical lines the subject exercised his legs on the ergometer. This caused his oxygen consumption to rise to an average of 690 ml/min above his resting oxygen consumption. Blood flow on the normal side fell slightly whereas on the nerve blocked side, the flow rose dramatically; this rise was presumably the passive consequence of the rise in arterial pressure. The difference between the responses on the two sides must have been due to the action of

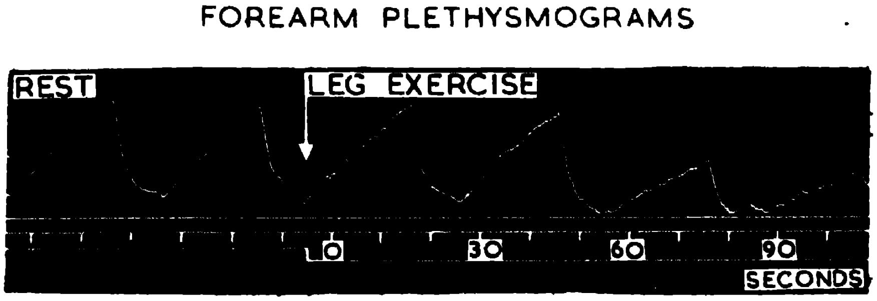


Fig. 4. Plethysmographic records of forearm blood flow before and at the beginning of a period of leg exercise. (Blair, Glover and Roddie, 1961*a*.)

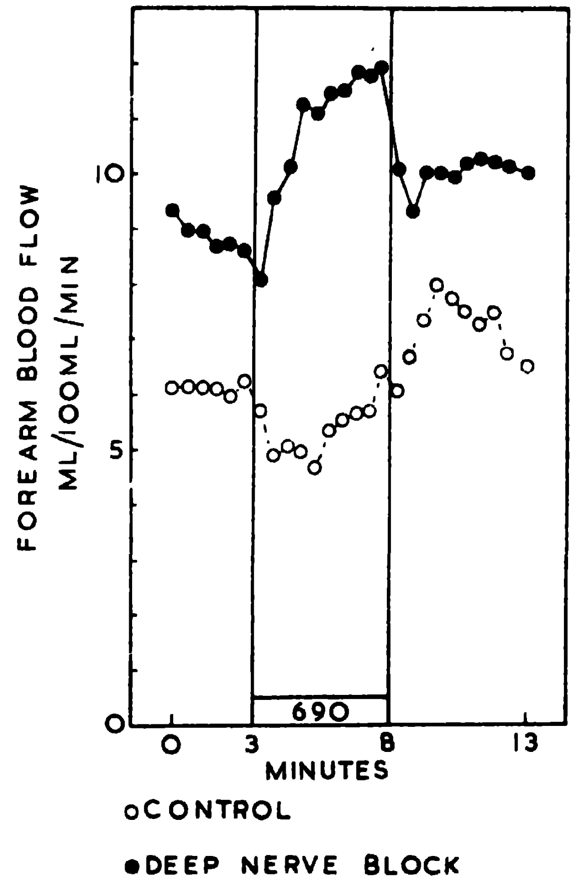


Fig. 5. The effect of blocking the deep nerves to the forearm with local anaesthetic on the blood flow changes during leg exercise. Average of 10 runs on three subjects. Control forearm (○); nerve blocked forearm(●). The rectangle represents the period of leg exercise and the number within the rectangle is the average increase in oxygen consumption above the resting level during exercise in ml/min. (Blair, Glover and Roddie, 1961*a*.)

vasomotor nerves. By their activity in the normal forearm they not only prevented blood flow from rising in the face of the increase in arterial pressure but actually reduced flow below the resting level. This was accomplished by producing a considerable rise in vascular resistance in the forearm. Blocking only the cutaneous nerves to the forearm had no effect on the vasoconstrictor response; this suggested that the nervous vasoconstriction was occurring deep to skin, i.e. in the muscle. Treating the forearm with bretylium tosylate, which blocks the activity of adrenergic vasoconstrictor fibres, blocked the vasoconstrictor response to leg exercise; this suggested that the nerves responsible for the nervous vasoconstriction in muscle were vasoconstrictor fibres. The vasoconstrictor response in the arm to exercise of the legs was related to the severity of the exercise but the location and the nature of the receptors concerned in this reflex are not known.

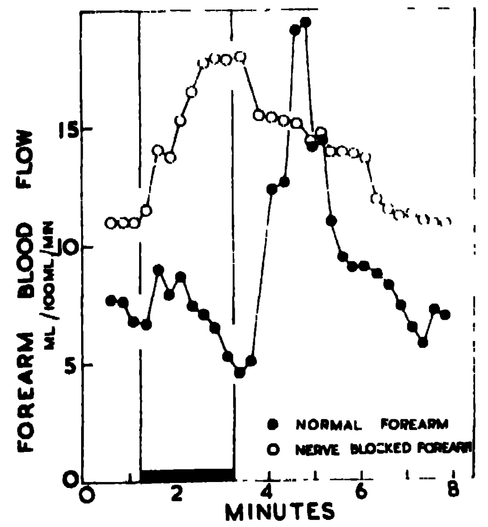


FIG. 6. A comparison of the changes in blood flow in a normal and in a nerve blocked forearm during the inhalation of 12·5 per cent CO_2 in O_2. (Blair, Glover, McArdle and Roddie, 1960.)

Inhalation of high concentrations of carbon dioxide also results in an increase in vasoconstrictor tone in muscle (Blair, Glover, McArdle and Roddie, 1960). Fig. 6 shows an experiment where the response to the inhalation of 12·5 per cent carbon dioxide in oxygen was compared in a normal and in a nerve blocked forearm. During inhalation the flow on the normal side fell slightly whereas that on the nerve blocked side rose; the rise was presumably due to the increased arterial pressure and perhaps a direct dilator effect of carbon dioxide on the forearm blood vessels. It would seem that vasomotor fibres normally increase resistance to blood flow in the forearm during inhalation of high concentrations of carbon dioxide. Resistance does not increase during inhalation in a forearm treated with bretylium tosylate; this suggests that vasoconstrictor fibres are responsible for the increase in resistance.

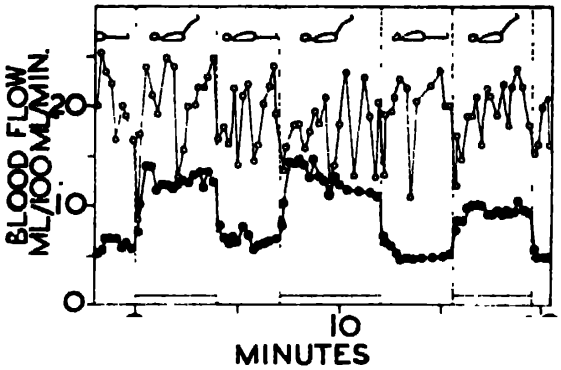


FIG. 7. The effect of passively raising the legs on forearm (●) and hand (○) blood flow. (Roddie and Shepherd, 1956.)

Stimuli which shift volumes of blood from one part of the body to another also cause changes in the activity of vasoconstrictor fibres. One of these stimuli is change of posture (Bridgen, Howarth and Sharpey-Schafer, 1950; Roddie and Shepherd, 1956). When the legs of a recumbent subject are passively raised, forearm blood flow increases whereas hand blood flow shows no comparable change (Fig. 7). The vasodilatation in the forearm depends on the integrity of the sympathetic nerve supply. Fig. 8 shows the result of an experiment where the effect of leg raising on forearm blood flow was compared before and several weeks after bilateral cervical sympathectomy. The vasodilatation induced by leg raising was abolished after sympathectomy. Bretylium tosylate also blocked the vasodilatation suggesting that the vasodilatation was brought about by release of vasoconstrictor tone. The release of vasoconstrictor tone occurred in muscle and not in skin. This is not surprising since, as I mentioned earlier, vasoconstrictor tone in skin is normally fully released when the subject is comfortably warm.

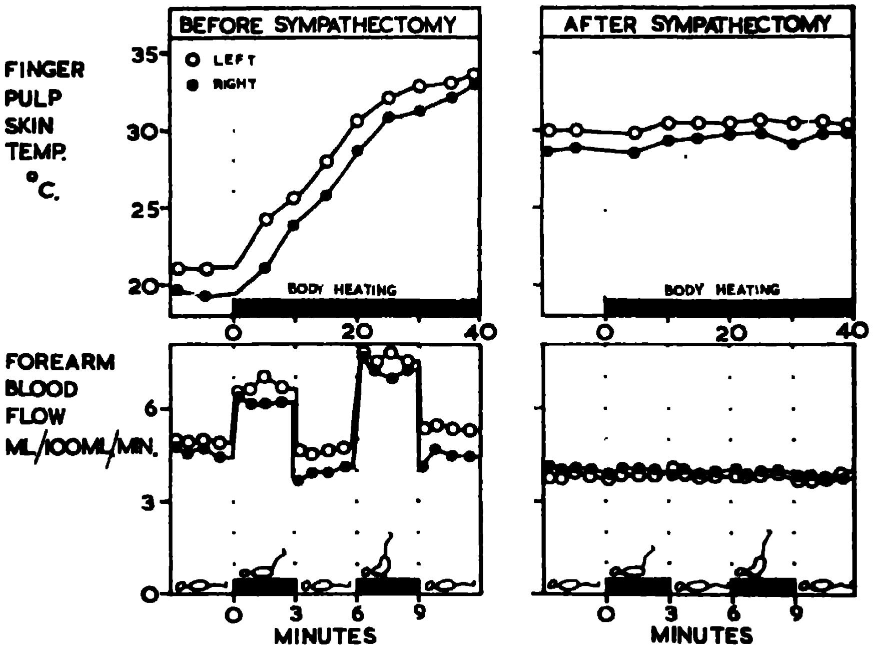


FIG. 8. The failure of forearm blood flow to increase in response to raising the legs following sympathetic neurotomy. *Left panels*: before operation. *Right panels*: after bilateral cervical sympathectomy. ○ left forearm; ● right forearm. (Roddie and Shepherd, 1956.)

Though the location and nature of the receptor concerned in this reflex are also unknown, the response did seem to depend on the shift of blood to the chest. Fig. 9 shows an experiment

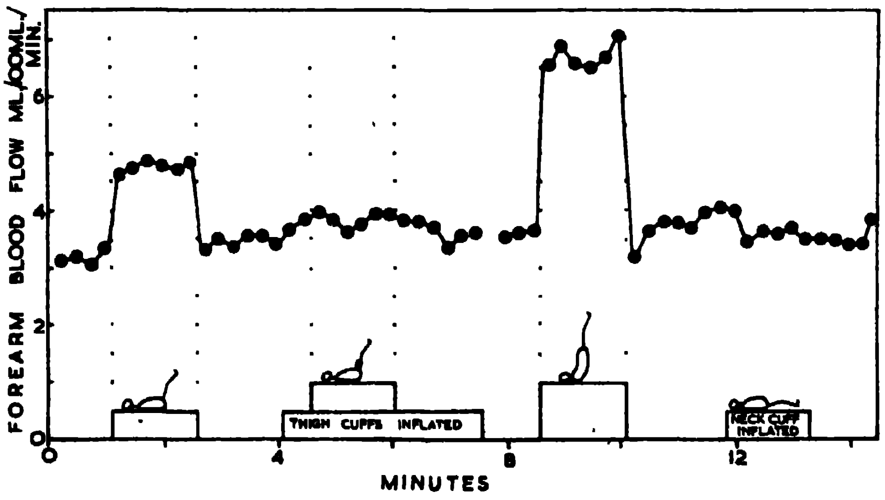


FIG. 9. The effect on forearm blood flow of (*a*) passively raising the legs, (*b*) passively raising the legs with the circulation to the legs occluded with a pneumatic cuff, (*c*) passively raising the legs and lower trunk, and (*d*) inflating a venous occlusion cuff on the neck. (Roddie, Shepherd and Whelan, 1957.)

where blood flow was measured in one forearm. During the period between the first pair of dotted lines there was the normal increase in flow which occurs when the legs were raised. When cuffs were inflated to 200 mm Hg around the thighs, raising the legs did not increase forearm blood flow. This suggested that the response was dependent on the shift of blood from the legs. Raising the legs and the trunk caused an even greater increase in flow than raising the legs alone. This suggested that the response did not depend on the pooling of blood in the splanchnic area. Increasing the blood volume in the head and neck by applying a venous occlusion cuff to the neck did not cause a vasodilatation in the forearm; the response did not therefore depend on the pooling of blood to the head. These experiments suggested that the response to raising the legs depended on the shift of blood to the intrathoracic vascular bed.

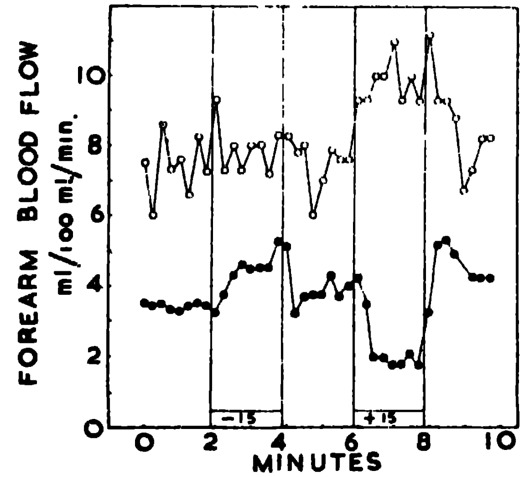


FIG. 10. Blood flow in normal (●) and nerve blocked (○) forearms during pressure breathing. During the two periods indicated, the subject breathed air at 15 mm Hg below and 15 mm Hg above atmospheric pressure. (Blair, Glover and Kidd, 1959.)

Tilting a subject from the horizontal into the foot down position, which tends to shift blood towards the lower extremities, results in a nervous vasoconstriction in the forearm (Bridgen *et al.*, 1950) and the degree of vasoconstriction is related to degree of tilt (Glover and Hutchison, 1964).

Positive and negative pressure breathing provide stimuli which tend to shift blood. Negative pressure breathing, which tends to increase intrathoracic blood volume, results in a nervously mediated vasodilatation in the forearm (Blair, Glover and Kidd, 1959). Positive pressure breathing, which tends to reduce the intrathoracic blood volume, results in a nervously mediated vasoconstriction (Fig. 10). The Valsalva manoeuvre which accomplishes the same shift in blood away from the chest also causes vasoconstriction in muscle (Roddie, Shepherd and Whelan, 1958).

Recently Brown, Goei, Greenfield and Plassaras (1964) have shifted blood volumes in the body by applying negative pressures to the legs and lower trunk. On applying suction which tends to shift blood towards the legs there is vasoconstriction in the forearm (Fig. 11). Releasing the suction which allows the trapped blood to return to the trunk caused vasodilatation. Similar changes were seen when blood trapped in the legs by venous tourniquets was released (Roddie and Shepherd, 1956). These effects are blocked by treatment of the arm with bretylium tosylate suggesting that vasoconstrictor fibres are responsible for the changes.

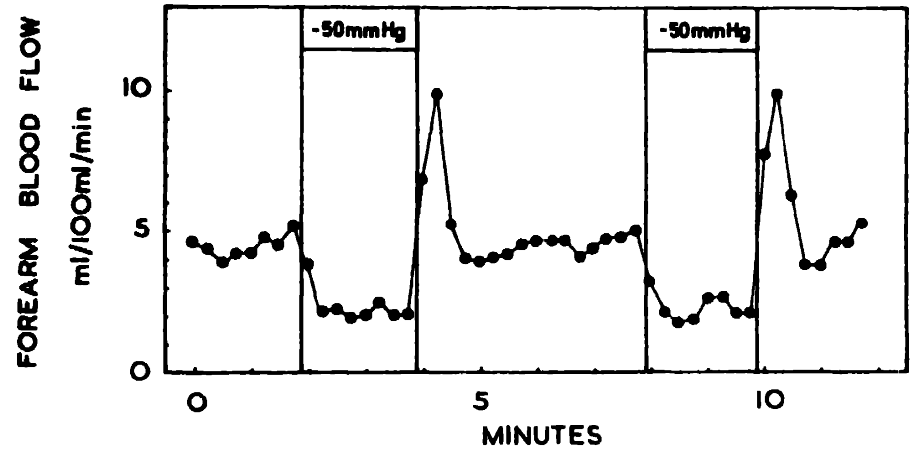


FIG. 11. The effect of applying a pressure of −50 mm Hg to the legs and lower trunk on forearm blood flow. (Glover and Hutchison, 1964.)

So here we have a reflex system which is very sensitive to changes in the distribution of blood in the body. The receptors concerned in the reflex are not known but they do not seem to lie in the high pressure system (Roddie and Shepherd, 1958).

In summary, the blood vessels in the skeletal muscle in the limbs are normally held under vasoconstrictor tone. The fibres are not concerned in temperature regulation and can function independently of the vasoconstrictor fibres to skin. Their activity is adjusted in response to a variety of stimuli such as muscular exercise, blood volume shifts and changes in gas tensions in the blood. However the receptors concerned in these reflexes have not yet been identified.

(2) *Vasodilator fibres*

In the cat and the dog a system of sympathetic vasomotor fibres has been described, whose stimulation results in a powerful vasodilatation in muscle blood vessels (Uvnas, 1954). They seem to act by releasing acetylcholine at the nerve terminals since their effect can be blocked by atropine. The central connexions and efferent distributions of these fibres was extensively studied but the nature of the stimulus which normally excites them remains obscure. The fibres and their central connexions were not involved in baro- or chemo-receptor reflexes in dogs and cats (Folkow and Gernandt, 1952; Frumin, Ngai and Wang, 1952; Lindgren and Uvnas, 1954). A clue to their physiological stimulus was provided by Abrahams and Hilton (1958). They found that stimulation by electrodes of places in the brain, which excited vasodilator fibres to muscle in anaesthetized cats, provoked a 'flight or fight' reaction in conscious cats. This suggested that vasodilator fibre discharge might be associated with flight or fight reactions.

The first evidence that human skeletal muscle had a vasodilator nerve supply was provided by Barcroft and Edholm (1945). They found that forearm blood flow increased during fainting induced by trapping blood in the limbs with venous tourniquets and by venesection. Since the skin appeared pale, it seemed reasonable to suppose that the increase was occurring deep to skin, that is, in the skeletal muscle. Since arterial pressure was falling rapidly at the time of the faint, the increase in flow was clearly due to vasodilatation. Blocking the deep nerves to the forearm changed this picture. During the faint, the blood flow in nerve blocked forearms fell, passively following the changes in arterial pressure. This suggested that the vasodilatation in the normal forearm was mediated by vasomotor nerves. The average blood flow at the height of a faint in nine normal forearms was higher than the average blood flow in six nerve blocked forearms. In other words the blood flow in the normal forearms during fainting rose to a level which could not readily be explained by release of vasoconstrictor tone alone. It was concluded that the vasodilatation in the normal forearm was due to activity of vasodilator fibres, rather than to release of vasoconstrictor tone.

Certain forms of emotional stress may also cause reflex excitation of these vasodilator fibres. Harassing a subject by giving him difficult mental arithmetic to perform has been shown by Abramson and Ferris, 1940; Wilkins and Eichna, 1941; Golenhofen and Hildebrandt, 1957; Brod, Fencl, Hejl and Jirka, 1959; and Barcroft, Brod, Hejl, Hirsjarvi and Kitchin, 1960 to cause vasodilatation in the muscular parts of the limbs. Wilkins and Eichna (1941) found that sympathetic fibres contributed to this response.

It is very difficult to frighten subjects deliberately under laboratory conditions but, working on the idea that the vasodilator system of fibres to muscle might be excited in association with flight or fright reactions, Blair, Glover, Greenfield and I (1959) attempted to induce fear in subjects. An experiment of this type is illustrated in Fig. 12. We led the subject, a medical student, to believe that he was losing blood very rapidly from an arterial puncture site, that his blood pressure was falling rapidly, and that his state of health was very precarious. This was all done by whispered conversation so that the subject could just hear. Though some of the people conducting the experiment made an effort to stop the experiment, others insisted that it must go on. The subject became quite worried and alarmed about this. His forearm blood flow rose to 50 ml/100 ml/min, a level as high as that seen after severe exercise of the forearm muscles. We then apologized and reassured the subject, explained the hoax, and his blood flow in the forearm fell very rapidly to the control level. During stress he felt ill. He complained of throbbing in the head and his arm was very sore where the blood was supposedly leaking. Though emotional stress is known to result in the release of adrenaline, this response did not show many of the characteristics of an adrenaline infusion. The rate of onset of vasodilatation and the rapidity of the fall were too great. The blood pressure response and the response of hand blood flow were not typical of adrenaline infusions. One would expect a much greater pressor response and vasoconstriction of the hand with doses of adrenaline that would cause such a large vasodilatation in the forearm.

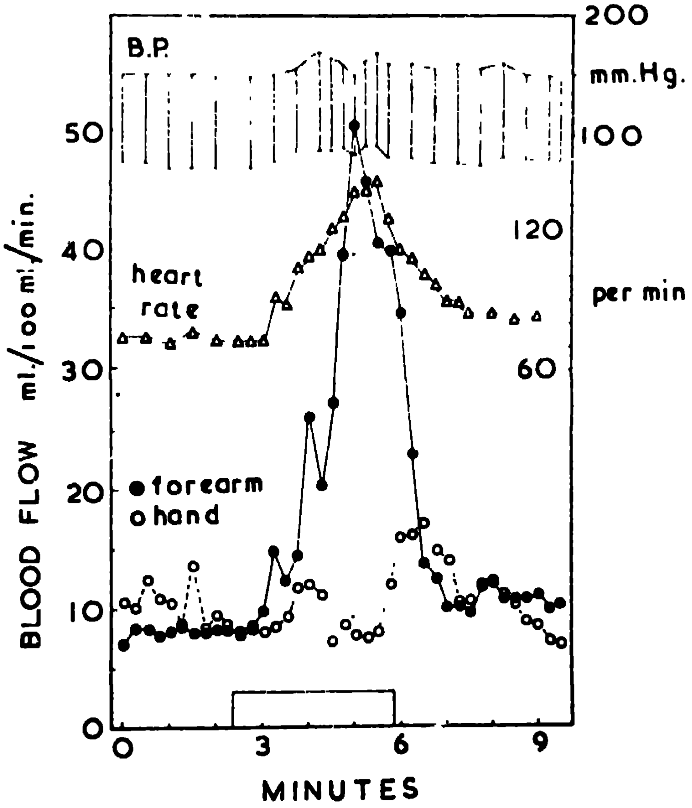


FIG. 12. Effect of severe emotional stress on arterial pressure, heart rate, forearm blood flow and hand blood flow. During the time represented by the rectangle, it was suggested to the subject that he was suffering from severe blood loss. (Blair, Glover, Greenfield and Roddie, 1959.)

Other experiments suggested that this type of vasodilatation occurred in muscle and not in skin, during the dilatation the oxygen saturation of venous blood from muscle rose whereas that from skin did not. In a subject who had had a unilateral sympathectomy the vasodilatation in response to fear was much less marked in the sympathectomized forearm; this showed that sympathetic nerves were concerned in the response. The response was reduced, though not abolished, by treating the forearm with atropine; this suggested that the sympathetic fibres were cholinergic. When all the evidence was summed it was concluded that sympathetic vasodilator fibres contribute to the vasodilatation in forearm muscle during certain varieties of emotional stress. The value to the organism of this redistribution of blood flow in states of stress is not at all clear. In the preparation of the organism for the violent exercise it would raise the internal temperature of the muscles in the limbs and permit a high cardiac output without much rise in arterial blood pressure. However, strenuous exercise can be performed very adequately by individuals whose limbs have been sympathectomized (Barcroft, 1962).

CONCLUSION

There are several systems of sympathetic fibres supplying the resistance blood vessels in the different tissues of the limbs. These systems can be independently activated in different circulatory reflexes.

Muscle blood vessels have both a constrictor and a dilator innervation. The constrictor fibres normally maintain a vasoconstrictor tone and this is varied reflexly in response to a variety of stimuli such as muscular exercise, alteration of the distribution of blood volume in the body and alteration of the level of carbon dioxide in the arterial blood. The vasodilator fibres to muscle are not active at rest but contribute to the vasodilatation in muscle during emotional stress.

The pattern of innervation of skin blood vessels is different in different parts of the limbs. In the hands and feet there seems to be an exclusively vasoconstrictor innervation. The considerable vasoconstrictor tone exerted by these fibres is reflexly varied in temperature regulation. In the skin of the more proximal parts of the limbs both a nervous vasoconstrictor and a nervous vasodilator mechanism are present though neither of these is active when the subject is comfortably warm. Cooling the body reduces blood flow by causing reflex activation of vasoconstrictor fibres and heating the body raises blood flow by the reflex excitation of nerves whose activity results in vasodilatation.

REFERENCES

ABRAHAMS, V. C. and HILTON, S. M. (1958) *J. Physiol.*, **140**, 16 P.

ABRAMSON, D. I. and FERRIS, E. B. (1940) *Amer. Heart J.*, **19**, 541.

ADSON, A. W. and BROWN, G. E. (1925) *Amer. J. med. Sci.*, **170**, 232.

BARCROFT, H. (1962) *Handbook of Physiology.* Circulation II, p. 1369. Baltimore. Williams and Wilkins Co.

BARCROFT, H., BOCK, K. D., HENSEL, H. and KITCHIN, A. H. (1955) *Pflug. Arch. ges. Physiol.*, **261**, 199.

BARCROFT, H., BONNAR, W. McK., EDHOLM, O. G. and EFFRON, A. S. (1943) *J. Physiol.*, **102**, 21.

BARCROFT, H., BROD, J., HEJL, Z., HIRSJARVI, E. A. and KITCHIN, A. H. (1960) *Clin. Sci.*, **19**, 577.

BARCROFT, H. and EDHOLM, O. G. (1945) *J. Physiol.*, **104**, 161.

BERNARD, C. (1852) *C.R. Soc. Biol.*, **4**, 168.

BLAIR, D. A., GLOVER, W. E., GREENFIELD, A. D. M. and RODDIE, I. C. (1959) *J. Physiol.*, **148**, 633.

BLAIR, D. A., GLOVER, W. E. and KIDD, B. S. L. (1959) *Clin. Sci.*, **18**, 9.

BLAIR, D. A. GLOVER, W. E., KIDD, B. S. L. and RODDIE, I. C. (1960) *Brit. J. Pharmacol.*, **15**, 466.

BLAIR, D. A., GLOVER, W. E., McARDLE, L. and RODDIE, I. C. (1960) *Clin. Sci.*, **19**, 407.

BLAIR, D. A., GLOVER, W. E. and RODDIE, I. C. (1960) *J. Physiol.*, **153**, 232.

BLAIR, D. A., GLOVER, W. E. and RODDIE, I. C. (1961) *J. appl. Physiol.*, **16** 119.

BLAIR, D. A., GLOVER, W. E. and RODDIE, I. C. (1961*a*) *Circulation Res.*, **9**, 264.

BRIGDEN, W., HOWARTH, S. and SHARPEY-SCHAFER, E. P. (1950) *Clin. Sci.*, **9**, 79.

BROD, J., FENCL, V., HEJL, Z. and JIRKA, J. (1959) *Clin. Sci.*, **18**, 269.

BROWN, ELLEN, GOEI. J. S., GREENFIELD, A. D. M. and PLASSARAS, G. (1964) *J. Physiol.*, **170**, 21 P.

EDHOLM, O. G., FOX, R. H. and MACPHERSON, R. K. (1956) *J. Physiol.*, **134**, 612.

EDHOLM, O. G., FOX, R. H. and MACPHERSON, R. K. (1957) *J. Physiol.*, **139**, 455.

FOLKOW, B. and GERNANDT, B. E. (1952) *Amer. J. Physiol.*, **169**, 622.

FOX, R. H., GOLDSMITH, R., KIDD, D. J. and LEWIS, H. E. (1962) *J. Physiol.*, **161**, 298.

FOX, R. H. and HILTON, S. M. (1958) *J. Physiol.*, **142**, 219.

FRUMIN, M. J., NGAI, S. H. and WANG, S. C. (1952) *Fed. Proc.*, **11**, 51.

GLOVER, W. E. and HUTCHISON, K. J. (1964) *J. Physiol.* In press.

GOLENHOFEN, K. and HILDEBRANDT, G. (1957) *Pflug. Arch. ges. Physiol.*, **263**, 637.

GRANT, R. T. and HOLLING, H. E. (1938) *Clin. Sci.*, **3**, 273.

LINDGREN. P. and UVNAS, B. (1954) *Amer. J. Physiol.*, **176**, 68.

McArdle, L. and Roddie, I. C. (1958) *Brit. J. Anaesth.*, **30**, 358.
Roddie, I. C. and Shepherd, J. T. (1956) *Clin. Sci.*, **15**, 433.
Roddie, I. C. and Shepherd, J. T. (1958) *Lancet*, **i**, 493.
Roddie, I. C., Shepherd, J. T. and Whelan, R. F. (1956) *J. Physiol.*, **134**, 612.
Roddie, I. C., Shepherd, J. T. and Whelan, R. F. (1957) *J. Physiol.*, **136**, 489.
Roddie, I. C., Shepherd, J. T. and Whelan, R. F. (1958) *Circulation Res.*, **6**, 232.
Uvnas, B. (1954) *Physiol. Rev.*, **34**, 608.
Walker, A. J., Lynn, R. B. and Barcroft, H. (1950) *St. Thomas's Hosp. Rept.*, **6**, 18.
Wilkins, R. W. and Eichna, L. W. (1941) *Johns Hopk. Hosp. Bull.*, **68**, 245.

XVII

Distribution of Pulmonary Blood Flow

JOHN B. WEST

INTEREST in the topographical distribution of pulmonary blood flow is both old and new; old because as long ago as 1887, Johannes Orth suggested that the weight of the column of blood in the lung might cause anaemia at the apex, and new because it has only been with the introduction of radioactive gases that it has become possible to measure the regional distribution of blood flow. One technique is shown in Fig. 1A. The subject is seated and pairs of scintillation counters are arranged over upper and lower zones of the lung. He takes a single breath of air containing carbon dioxide labelled with oxygen-15 and holds his breath for some 15 seconds. When the radioactive gas enters the counting field, there is an increase in counting rate which depends on the regional ventilation of the lung and its volume. During the subsequent breath-holding period, radioactivity can only be removed by the blood flow, so that the slope of the tracing or 'clearance rate' is taken as a measure of regional perfusion (West and Dollery, 1960).

Another technique is shown in Fig. 1B. This time the relatively insoluble xenon-133 is used (Ball, Stewart, Newsham and Bates, 1962). The gas is dissolved in saline and injected into a peripheral vein. When it reaches the lung, it is evolved into alveolar gas because of its low solubility and it remains there during a short breath-holding period. The distribution of intrapulmonary radioactivity can then be detected either by multiple external counters or, conveniently, by scanning each lung from bottom to top with a pair of counters. This scan is then compared with another made after mixing the xenon uniformly throughout the alveolar gas by rebreathing. In this way, blood flow per unit alveolar volume is derived.

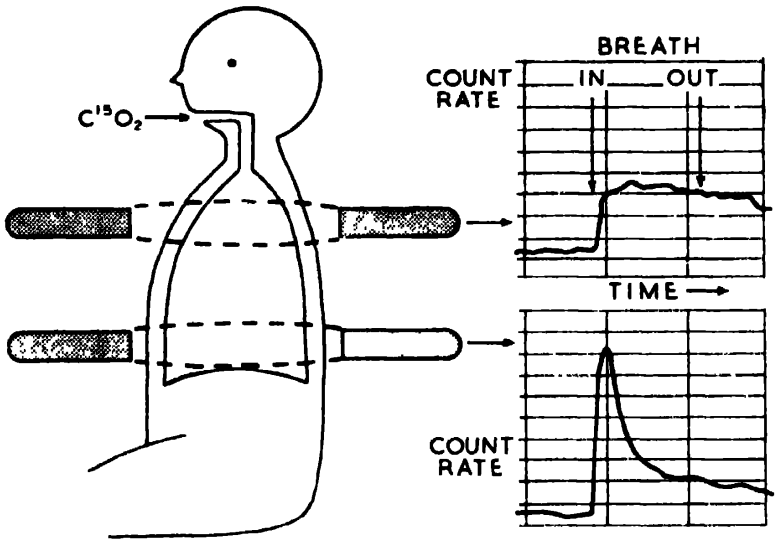


FIG. 1A. Technique of measuring the regional distribution of pulmonary blood flow and ventilation using oxygen-labelled carbon dioxide. The subject takes a single breath and the increase in counting rate reflects the regional ventilation, while the 'clearance rate' of the radioactive gas during breath-holding measures the regional blood flow. (From *Brit. Med. Bull.*)

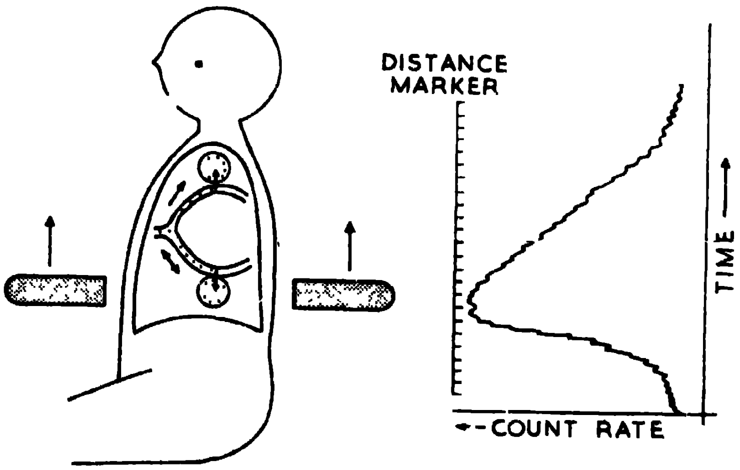


FIG. 1B. Technique of measuring regional blood flow using radioactive xenon. The gas is dissolved in saline and injected into a peripheral vein. When it reaches the lung it is evolved into alveolar gas and the lung is scanned during breath-holding from bottom to top by two opposed counters. (From *Recent Advances in Respiratory Physiology*.)

NORMAL DISTRIBUTION OF BLOOD FLOW

Fig. 2 shows the distribution of blood flow in the normal upright lung. It can be seen that blood flow decreases fairly steadily from the bottom to the top there being virtually no flow at the apex. These results were obtained with radioactive carbon dioxide and the general pattern has been confirmed with xenon

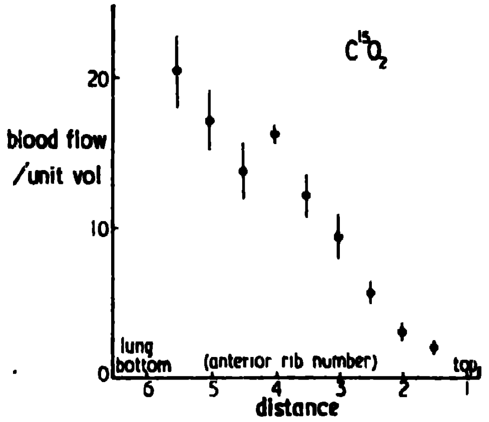


FIG. 2. Normal distribution of blood flow in the upright human lung as measured with radioactive carbon dioxide. Means and standard errors. Note the steady decrease from the bottom to the top of the lung; flow is almost nil at the apex. (Modified from *J. app. Physiol.*)

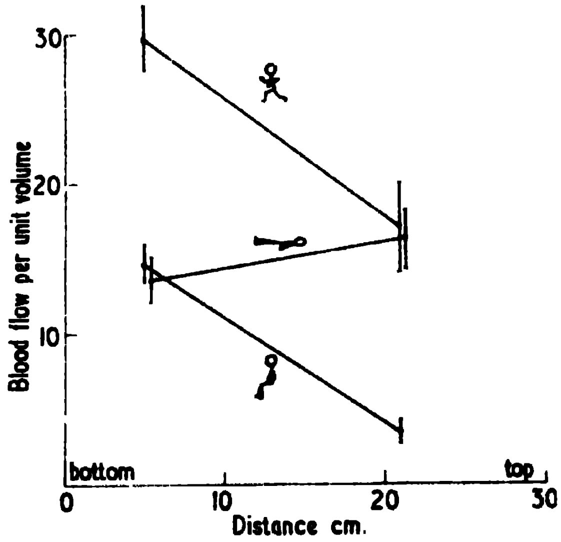


FIG. 3. Effects of change of posture and exercise on the distribution of blood flow. Means and standard errors. Note that in the supine position, upper and lower zone blood flows are virtually the same, and on exercise in the upright position, upper zone blood flow rises more than lower zone flow.

(Bryan, Bentivoglio, Beerel, MacLeish, Zidulka and Bates, 1964). This normal distribution is affected by change of posture and exercise. Fig. 3 shows that when the subject lies supine, apical blood flow increases until it is approximately the same as basal flow. On mild exercise, apical blood flow increases more than the basal flow so that the distribution becomes more uniform. Pathology also affects the normal distribution of blood flow. Left-to-right intracardiac shunts which increase the pulmonary blood flow produce the same pattern as exercise. Pulmonary hypertension not associated with an increased pulmonary blood flow results in a more even distribution than in the normal lung. Diseases in which the pulmonary venous pressure is moderately raised cause an even distribution while higher venous pressures may be associated with an inversion of the normal pattern of flow so that apical blood flow exceeds basal flow.

CAUSE OF THE NORMAL DISTRIBUTION

In an effort to understand the role of the various pressures in the lung circulation in determining the pattern of blood flow, we have studied an isolated lung preparation (West, Dollery and Naimark, 1964). The left lung of a dog is removed, suspended in a box, ventilated with negative pressure and perfused with a non-pulsatile flow of venous blood from a second dog. Various features of this preparation such as its ability to exchange gas, the compliance of the lung, the pulmonary vascular resistance, and the rise in this resistance during hypoxia, demonstrate its normal physiological behaviour. Fig. 4A shows the distribution of blood flow found using radioactive xenon in this preparation when the pulmonary arterial pressure was approximately normal. It can be seen that blood flow decreased steadily from the bottom to the top of the lung to virtually nil at the apex. The similarity of this distribution to that found in the upright human lung (Fig. 2) is striking.

Fig. 4B shows the distribution of blood flow in the same lung a few minutes later when the pulmonary arterial pressure was lowered by reducing the blood flow through the lung. Again blood flow decreased with distance up the lung, but on this occasion it fell to zero some two-thirds of the way up the lung so that an appreciable part of the apex was unperfused. It is possible to determine the level in the lung where pulmonary arterial pressure, as measured at the arterial cannula, equalled alveolar pressure. Here we are referring to the fact that the arterial pressure falls at the rate of 1 cm H_2O per cm of distance up the lung because of the hydrostatic effect. Fig. 4B shows good agreement between the no-flow level and the level at which arterial pressure equalled alveolar pressure and this was confirmed on many occasions.

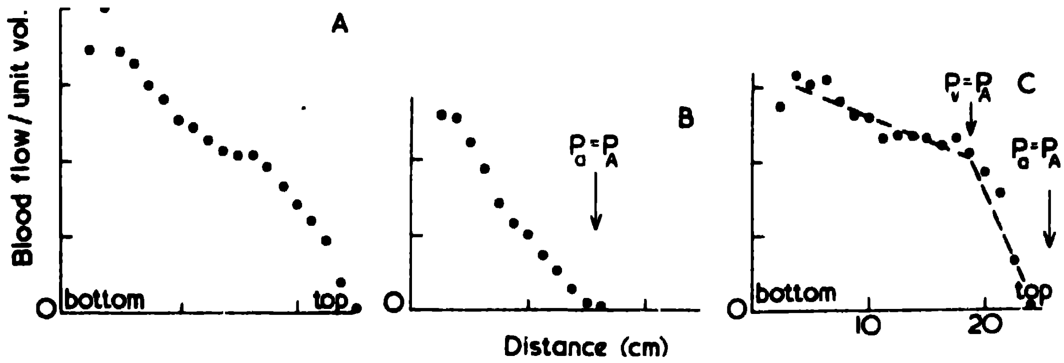


FIG. 4. Distribution of blood flow in the isolated dog lung under various conditions. A shows the flow distribution found with a normal pulmonary arterial pressure (P_a) and low venous pressure. Note the similarity to the distribution in the upright human lung shown in Fig. 2. B shows the effect of lowering the pulmonary arterial pressure. Note that there is now an unperfused zone at the top of the lung and that blood does not rise above the level at which arterial pressure equals alveolar pressure (P_A).

C shows the effect of raising the pulmonary venous pressure (P_V). The result was to make the flow distribution more even below the level at which venous pressure equals aveolar pressure. (From the *Handbook of Physiology*.)

Fig. 4C shows the effect on the distribution of blood flow of raising the pulmonary venous pressure which was kept low for Fig. 4A and B. It can be seen that the result was to make the distribution of blood flow in the lower part of the lung more uniform; instead of the distribution being approximately linear, an inflection is now seen and other evidence confirms that this occurs at the level where pulmonary venous pressure equals alveolar pressure.

Fig. 5 shows a scheme which satisfactorily accounts for the various distributions of blood flow which have been observed in the isolated dog lung. Many features of this have been previously proposed by Permutt, Bromberger-Barnea and Bane (1962), Banister and Torrance (1960) and others. The lung is divided into three zones by the relative magnitudes of the pulmonary arterial, venous and alveolar pressures.

In *Zone 1*, arterial pressure is less than alveolar pressure and there is no flow. This is presumably because collapsible vessels (possibly the pulmonary capillaries) are directly exposed to alveolar pressure.

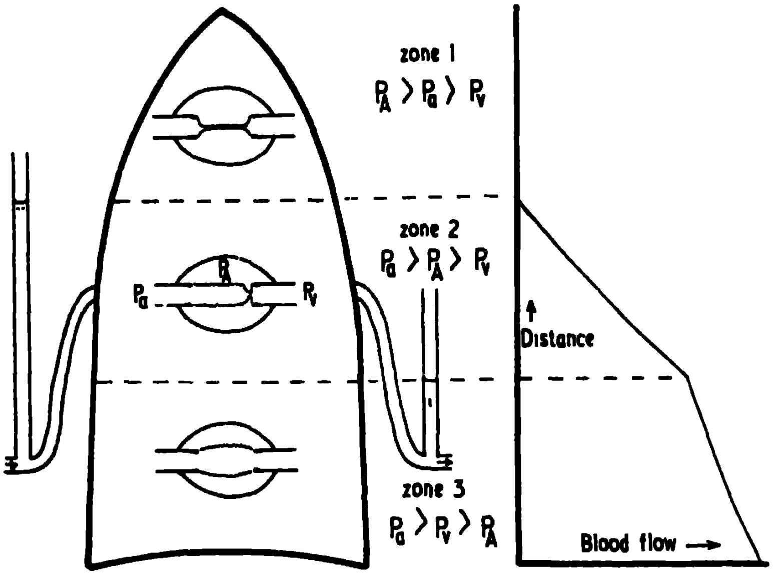


FIG. 5. Diagram to explain the topographical distribution of blood flow in the lung. The lung is divided into three zones by the relative magnitudes of the pulmonary arterial, venous and alveolar pressures. See text for details. (From *J. app. Physiol.*)

In *Zone 2*, arterial pressure exceeds alveolar pressure and alveolar exceeds venous pressure. Here the vessels behave like the resistor in a Starling heart-lung preparation which consists of a collapsible tube surrounded by a pressure chamber. When chamber pressure (alveolar) exceeds downstream pressure (venous), flow is determined not by the ordinary arterial-venous difference but by the arterial-alveolar difference. This is because a constriction develops in the downstream end of the collapsible tube until the pressure inside the tube equals the pressure outside (alveolar). Arterial pressure increases linearly down zone 2 while alveolar pressure is constant. Thus it is not surprising that blood flow is observed to increase steadily down this zone (Fig. 4A and B).

In *Zone 3*, arterial pressure exceeds venous pressure and now venous exceeds alveolar pressure. Flow is determined in the usual way by the arterial-venous difference, and since this is constant down the zone, it is not obvious at first why flow increases down this zone. However, since the pressure inside the vessels increases down the zone while the pressure outside (alveolar or pleural) is constant, the transmural pressure increases down the zone. Thus if parts of the vessels are distensible, their calibre will increase and their resistance to flow decrease. The increase in blood flow down zone 3 is therefore explained but it is not surprising that it is different from the increase in flow down zone 2 since these are determined by different mechanisms.

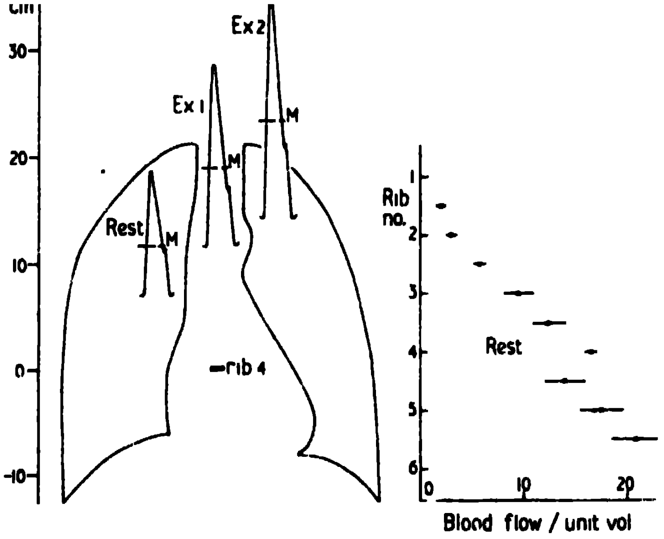


Fig. 6. Diagram of an upright human lung with the levels to which blood flow will rise during the cardiac cycle during rest and two levels of exercise (oxygen consumptions of about one and two litres per minute respectively). Data from Bevegard *et al.* (1960). M denotes mean pressure; reference level is insertion of 4th rib at sternum. Note that at rest, the apex of the lung will be totally unperfused, while some one third to one half of the lung will be unperfused during part of the cardiac cycle. On exercise, increasing amounts of lung will be perfused but even during moderately severe work, some of the lung will be unperfused some of the time. To the right is the distribution of blood flow as measured at rest with oxygen-labelled carbon dioxide (same data as Fig. 2). Agreement with the resting pulmonary artery pressure is reasonably good.

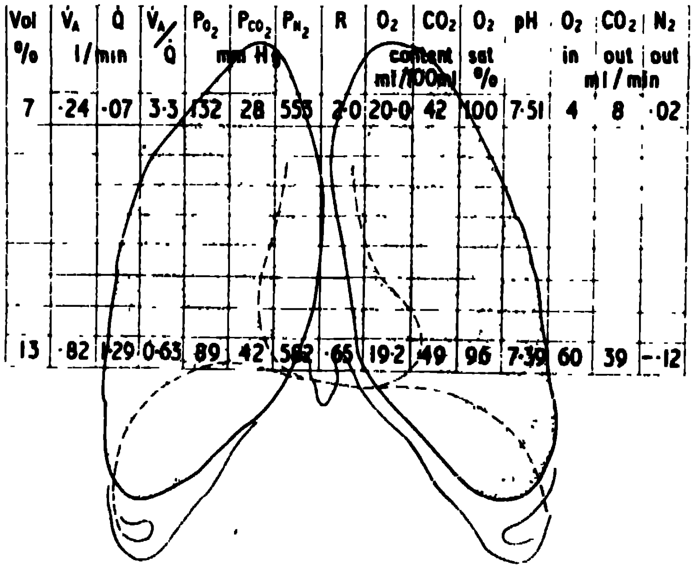
Application of this scheme to man is complicated by the pulsatile nature of the pulmonary arterial pressure. Fig. 6 shows pulmonary arterial pressures in man in the sitting position as measured by Bevegard, Holmgren and Jonsson (1960). If blood rises only to the level at which arterial pressure equals alveolar pressure, these measurements show that about one third of the lung is unperfused during part of the cardiac cycle at rest and that even on moderately severe exercise, not all the lung is perfused all the time. On the right side of the diagram, the distribution of blood flow found at rest with radioactive carbon dioxide is shown, and this agrees well with the measured pulmonary arterial pressure. It may be that these large pressure changes are not transmitted to the small vessels or that the vessels do not follow them faithfully so that the no-flow level is given by the mean pressure which is also shown.

The scheme of Fig. 5 accounts for most of the distributions of blood flow so far observed in the human lung. The fact that apical and basal flows become equal in the supine position follows from the evening out of the hydrostatic pressure differences. The more uniform distribution of blood flow on exercise compared with rest results from the increased pulmonary arterial pressure, and the same explanation applies to pathological conditions such as left-to-right intracardiac shunts in which pulmonary blood flow is increased. The fact that the distribution of blood flow becomes more uniform when the venous pressure is moderately raised is explained by the upward extension of zone 3 as well as the passive rise in pulmonary arterial pressure. However, the inverted distribution of blood flow seen in patients with a greatly raised venous pressure cannot be explained on the basis of Fig. 5 (see below).

REGIONAL GAS EXCHANGE

An interesting consequence of the uneven distribution of blood flow in the lung is the large regional differences in gas exchange which must be present. While blood flow decreases rapidly with distance up the lung, ventilation decreases much more slowly (West and Dollery, 1960; Ball *et al.*, 1962) with the result that the ratio of ventilation to blood flow increases strikingly up the lung.

This ventilation-perfusion ratio determines the gas exchange which occurs in any part of the lung and Fig. 7 shows the resulting differences which are calculated to exist in the normal upright lung (West, 1962). For these calculations, the lung was divided into nine imaginary horizontal slices corresponding to the nine counting positions originally used to collect the radioactive carbon dioxide data (Fig. 2). It was necessary to assume reasonable values for the cardiac output and the overall ventilation-perfusion ratio, and the approximate volume of the different slices was obtained by mensuration.



Vol %	$\dot{V}_A$ l/min	$\dot{Q}$ l/min	$\dot{V}_A/\dot{Q}$	P_{O_2} mm Hg	P_{CO_2} mm Hg	P_{N_2} mm Hg	R	O_2 content ml/100ml	CO_2 content ml/100ml	O_2 sat %	pH	O_2 in ml/min	CO_2 out ml/min	N_2 out ml/min
7	·24	·07	3·3	132	28	553	2·0	20·0	42	100	7·51	4	8	·02
13	·82	1·29	0·63	89	42	582	·65	19·2	49	96	7·39	60	39	–·12

FIG. 7. Effects of the topographical distribution of pulmonary blood flow and ventilation on regional gas exchange. In this diagram, the lung is divided into nine imaginary horizontal slices, but only the figures for the uppermost and lowermost slices are shown. The table shows relative lung volume (as percentage of total examined), ventilation ($\dot{V}_A$), blood flow ($\dot{Q}$), ventilation-perfusion ratio ($\dot{V}_A$: $\dot{Q}$), gas tensions (P_{O_2}, P_{CO_2} and P_{N_2}), respiratory exchange ratio (R), end-capillary blood values (contents of O_2 and CO_2, O_2 saturation, pH) and local gas exchange (O_2, CO_2 and N_2). (Modified from *J. app. Physiol.*)

Fig. 7 shows that the alveolar oxygen tension changes by more than 40 mm Hg down the lung and appreciable differences in carbon dioxide and nitrogen tension also occur. As a result, the gas composition of blood draining from the various levels is different and there is a surprisingly large change in pH from the top to the bottom. Note that the uppermost slice transfers only 4 ml of oxygen per minute compared with the 60 ml per minute of the lowermost slice. Thus the apex of the upright lung at rest takes little share of the overall gas exchange though its role is greater on exercise when blood flow is more evenly distributed. Nitrogen actually passes out of the blood into the gas at the apex and in the reverse direction at the base because of the regional differences in ventilation-perfusion ratio, so that there is a nitrogen cycle within the lung.

OVERALL GAS EXCHANGE

A remarkable feature of these large regional differences in gas tensions and contents within the lung is that they have so little effect on the overall gas exchange of the organ.

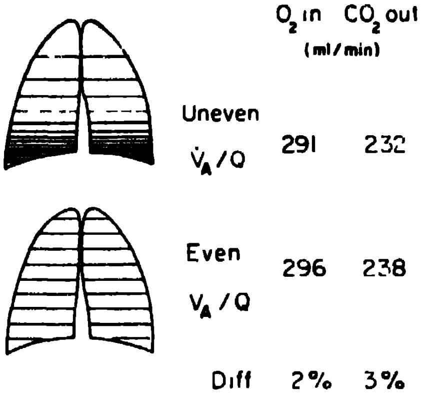
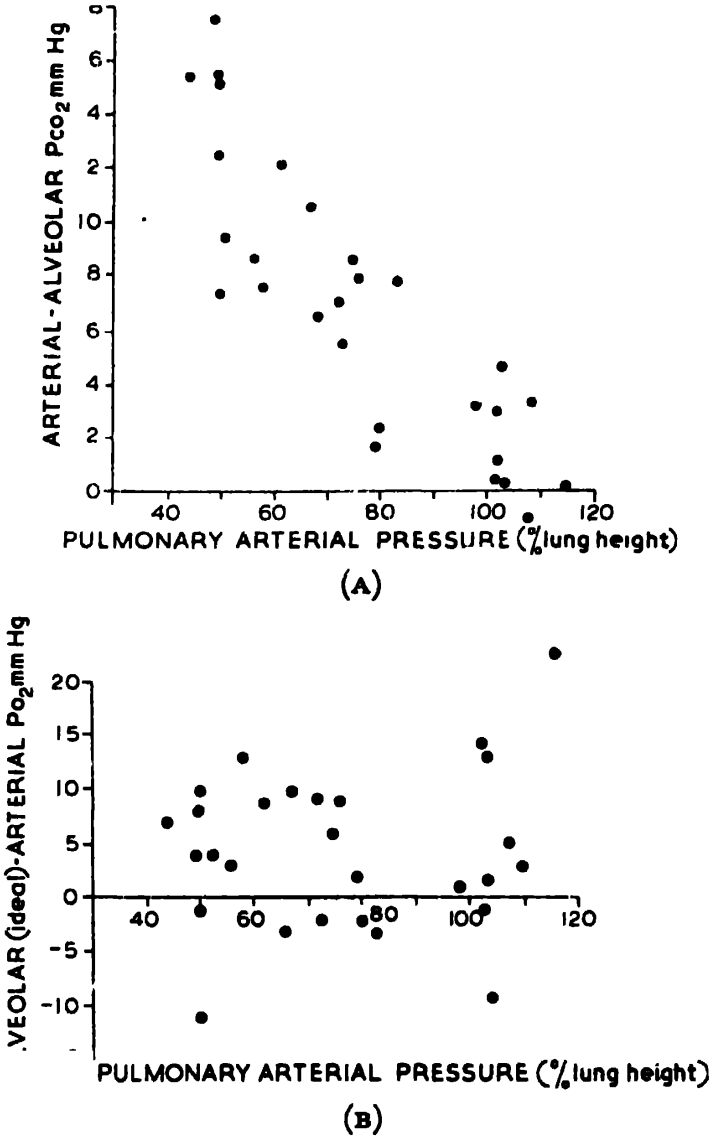


FIG. 8. Effect of the uneven distribution of blood flow and ventilation in the normal lung on overall gas exchange. The amounts of oxygen and carbon dioxide transferred by a model lung are compared before and after the imposition of ventilation-perfusion ratio inequality. Note that result of the uneven distribution is to reduce oxygen uptake and carbon dioxide output by only 2 and 3 per cent respectively.

Fig. 8 compares the amounts of oxygen and carbon dioxide transferred by two model lungs; both have the same total blood flow and ventilation and are supplied with the same venous blood and inspired air, but in one lung blood flow and ventilation are uniform, while in the other, they are unevenly distributed as in the normal upright lung. It can be seen that the result of imposing this ventilation-perfusion ratio inequality was to reduce the volume of oxygen transferred by the lung by 2 per cent, and the volume of carbon dioxide by 3 per cent. These are very small reductions which can easily be made up by small increases in total blood flow and ventilation. Thus it is clear that although the large inequalities of blood flow and ventilation which occur in the normal lung result in striking regional differences in gas exchange, they have a negligible effect on the overall gas exchange function of the lung which is what matters to the organism as a whole.

It is possible to calculate the gas tension differences which develop between mixed alveolar gas and arterial blood as a result of the topographical inequalities shown in Fig. 7. They are 4, 1 and 3 mm Hg for oxygen, carbon dioxide and nitrogen,



(A)

(B)

FIG. 9. Effect of reducing the pulmonary arterial pressure on gas exchange in the isolated lung. A shows that the arterial-alveolar difference for carbon dioxide rises rapidly as the arterial pressure is reduced implying an increasingly large physiologic dead space. B shows however that the ideal alveolar-arterial difference for oxygen is virtually unchanged by pulmonary hypotension and that there is little venous admixture. (From *J. app. Physiol*).

respectively, and these low values confirm the small impairment of overall gas exchange. The measured alveolar-arterial oxygen difference in the normal lung is about 10 mm Hg of which about 5 mm Hg must be accounted for by uneven distribution. The fact that the topographical inequality of blood flow and ventilation causes a calculated oxygen difference of 4 mm Hg suggests that this constitutes the bulk of the ventilation-perfusion ratio inequality in the normal lung.

While the uneven distribution of blood flow in the normal lung causes little impairment of overall gas exchange under normal conditions, there are circumstances where considerable disorder of function occurs. Thus if the pulmonary arterial pressure is lowered or the alveolar pressure is raised, the junction of zones 1 and 2 (Fig. 5) is depressed and increasing amounts of lung are unperfused. The resulting effects on gas exchange have been studied in the isolated lung by gradually reducing the pulmonary arterial pressure. Fig. 9A shows that as the arterial pressure was reduced below the level necessary to perfuse the whole lung (100 per cent on the horizontal axis), an increasing arterial-alveolar difference for carbon dioxide developed. This implies the formation of a large alveolar dead space and consequent interference with the ability of the lung to put out carbon dioxide. By contrast, Fig. 9B shows that the ideal alveolar-arterial difference for oxygen was not systematically changed as the arterial pressure was reduced. This means that the venous admixture effect was small and that the ventilation-perfusion ratio inequality interfered little with oxygen transfer by the lung (West and Jones, 1965).

Among the conditions in which this pattern of disordered function occurs are haemorrhage, shock, anaesthesia and positive-pressure breathing. In all these, increased physiologic dead spaces have been described (Gerst, Rattenborg and Holaday, 1959; Freeman and Nunn, 1963; Bitter and Rahn, 1956), while venous admixture effects are generally small unless atelectasis is also present. Exposure of the lung to high accelerations also causes a large alveolar dead space for the same reason though here there may be severe arterial hypoxaemia as well because of the marked inequality of ventilation (Barr, 1963; Wood, Nolan and Donald, 1963).

PRESSURE-FLOW RELATIONS

An important consequence of the uneven distribution of blood flow in the lung is that the pressure-flow relations of the lung are greatly affected by the pattern of blood flow present (West and Dollery, 1965). It is clear from Fig. 5 that the vascular resistances of zones 1, 2 and 3 are inherently different. The resistance of zone 1 is infinitely high because the vessels are closed; the resistance of zone 2 is moderately high because the vessels are constricted at one end by the alveolar pressure; and the resistance of zone 3 is the lowest because there the vessels are distended. Thus if for any reason zone 2 encroaches on zone 1 because the pulmonary arterial pressure is raised, the pulmonary vascular resistance falls; in the same way if zone 3 encroaches on zone 2 because the venous pressure is increased, the pulmonary vascular resistance will also fall. This fall in vascular resistance which occurs when either the pulmonary arterial or venous pressure is raised is a valuable feature for the only organ in the body which must accept the total cardiac output under a variety of conditions.

These changes are illustrated by gradually increasing the pulmonary arterial pressure in an isolated lung and observing the blood flow which results (Fig. 10A). It can be seen that as the arterial pressure was gradually raised from a low value, flow first increased slowly, then more rapidly as zone 2 extended upwards and more and more vessels opened. When the arterial pressure exceeded alveolar pressure at the top of the lung so that all the vessels were open, the relationship between flow and pressure was nearly linear. Fig. 10B shows the result of gradually raising pulmonary venous pressure constant while holding blood flow rate constant. Initially, changes in venous pressure had no effect on arterial pressure because all the vessels were in zone 2. (The fall in pulmonary vascular resistance at this time was not due to any change in the behaviour of the lung vessels but simply reflected the conventional formula for calculating pulmonary vascular resistance which contains venous pressure.) However, when venous pressure exceeded alveolar pressure at the bottom of the lung so that some vessels were in zone 3, arterial pressure began to rise; when arterial pressure exceeded

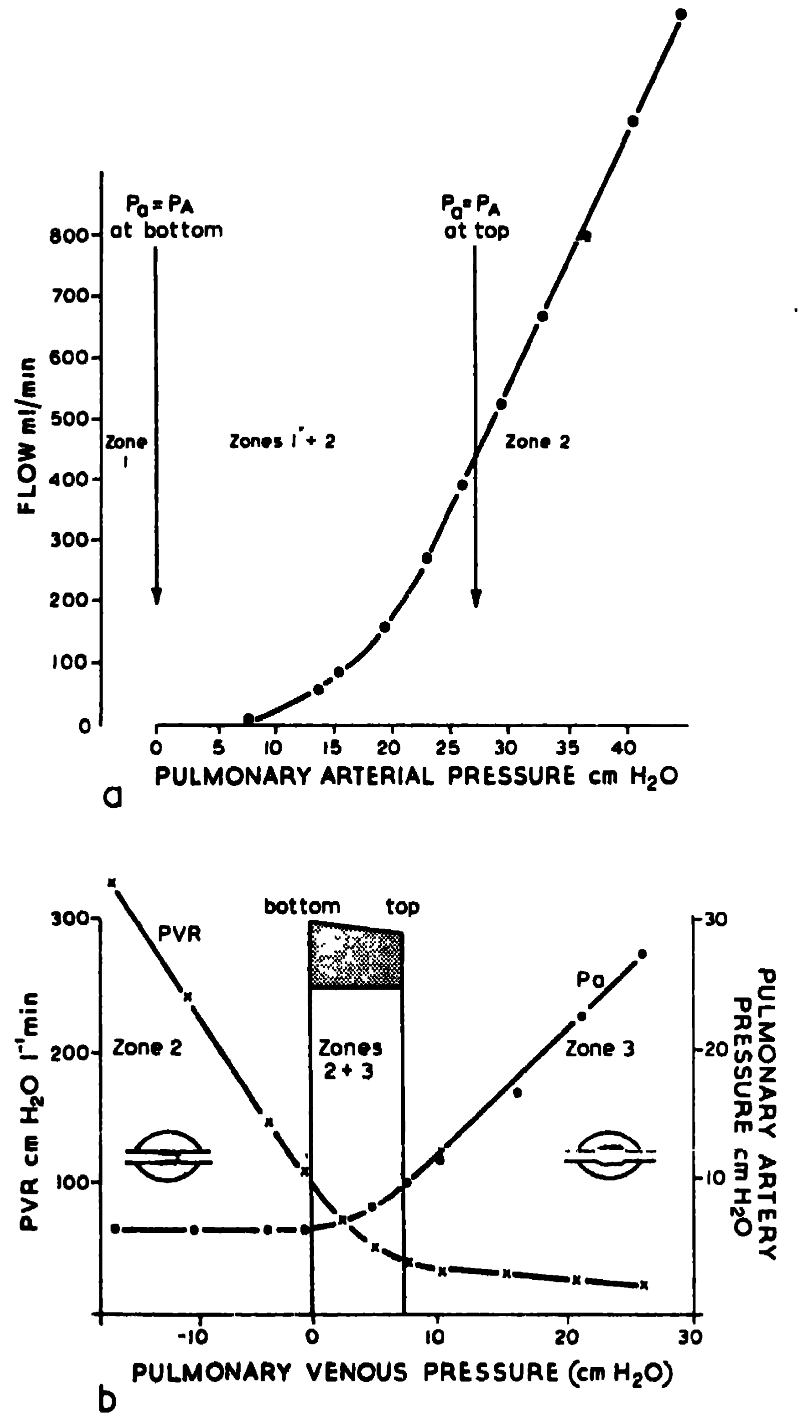


FIG. 10. Effect of the uneven distribution of pulmonary blood flow on the pressure flow relations of the lung. A shows the way flow changes as the pulmonary arterial pressure (P_a) is gradually raised. B shows the way arterial pressure and pulmonary vascular resistance (PVR) change when venous pressure is gradually increased from very low values. For explanation, see text. (From *J. app. Physiol.*)

alveolar pressure at the top of the lung, venous pressure rose linearly with arterial pressure and the steady slight fall in pulmonary vascular resistance was attributable to the increasing calibre of the vessels. Thus what appears at first sight to be a very complicated pattern of pressure-flow relations falls simply into three parts when the topographical distribution of blood flow is taken into account.

ABNORMAL DISTRIBUTION OF BLOOD FLOW CAUSED BY INTERSTITIAL OEDEMA

We have seen that hydrostatic pressure differences within the lung can apparently explain all the patterns of blood flow distribution described in man so far with one striking exception. Patients with severe mitral stenosis (Dollery and West, 1960) and acute left ventricular failure may develop an inversion of the normal pattern so that the blood flow to the upper zone exceeds that to the lower zone (Fig. 11A). It is not possible to account for this distribution on the basis of the scheme shown in Fig. 5.

It was therefore surprising to find a similar pattern of blood flow in the isolated lung under special conditions, namely when the pulmonary venous pressure was raised and the arterial-venous pressure difference was small (West, Dollery and Heard, 1965). Fig. 11B shows an example in a lung which had been perfused for about two hours and in which the venous pressure was 17 cm H_2O (referred to the hilum). It can be seen that most of the lower lobe is virtually unperfused implying a great increase in vascular resistance in this area.

There is good evidence that this increase in vascular resistance occurring in the dependent zone of the isolated lung was caused by interstitial perivascular oedema obstructing blood flow. Thus the effect generally became more marked as perfusion was continued and lungs which were becoming oedematous showed the effect well, though good examples were seen in lungs which gained little weight and had no intra-alveolar oedema. The increased vascular resistance was always seen in the dependent parts of the lung even if it was suspended in the inverted posture. The effect was most marked the smaller the arterial-venous pressure difference, and lowering the venous pressure always greatly increased the flow to the dependent zones even if total blood flow was kept constant. The increased vascular resistance in the dependent zone was unaffected by infusing isoprenaline or acetylcholine into the pulmonary artery, but a marked transient reduction in the effect followed the infusion of hypertonic urea solution into the pulmonary artery.

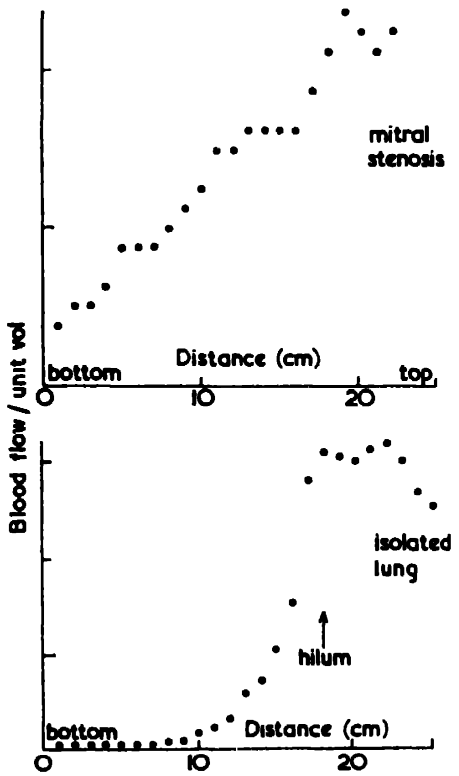


FIG. 11. A. Distribution of pulmonary blood flow in a 43-year-old housewife with severe mitral stenosis. Note the steady increase in blood flow from the bottom to the top of the lung which is the inverse of the normal pattern (compare Fig. 2). B. Example of the distribution of blood flow in an isolated lung with a raised pulmonary venous pressure. Note the striking reduction of blood flow (increased pulmonary vascular resistance) in the dependent part of the lung compared with the expected pattern (compare Fig. 4C). (From *The Lancet.*)

Supporting evidence was provided by histological sections which were taken while the lung was being perfused, by a rapid freezing method (Staub and Storey, 1962). The sections showed thick cuffs of perivascular oedema around both arteries and veins, vessels of diameters of 100 microns to several mm being affected. Similar histological changes have been described by Trowell (1943) in an isolated lung preparation.

At first sight it seems improbable that perivascular oedema could increase vascular resistance in such an airy, distensible organ as the lung, although in a solid structure like the kidney, the correlation is accepted (Hinshaw, Day and Carlson, 1959). However, Macklin (1946) and Howell, Permutt, Proctor and Riley (1961) have suggested that some of the pulmonary vessels are held open by a low perivascular pressure which is developed by inflation of the lung. This notion is supported by our observations that the expanded lung does not show a critical closing pressure (that is, a perfusing pressure at which no flow occurs) while the collapsed lung does have a critical closing pressure of about 7 cm H_2O. Thus a possible mechanism by which perivascular oedema could increase vascular resistance in the lung is shown in Fig. 12. Normally the vessels outside the influence of alveolar pressure have an inherent tendency to close because of tension in their walls but they are held open by the expanding action of the lung (Fig. 12A). In the collapsed lung, they show a critical closing pressure and an increased resistance (Fig. 12B).

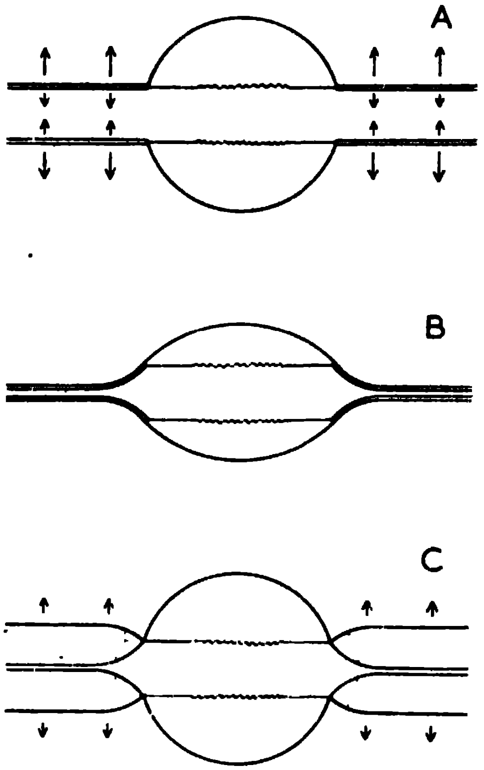


FIG. 12. Possible mechanism by which interstitial oedema causes an increased vascular resistance in the dependent zone of the isolated lung. For explanation, see text. (From *Circ. Res.*)

Finally interstitial oedema distends the potential perivascular space (Tocker and Langston, 1952) and isolates the contracting vessels from the usual expanding force (Fig. 12C).

Whether this mechanism ever operates in the human lung can only be a matter for speculation as yet but certainly enlargement of the perivascular space by distended lymphatics is known to occur in patients with high pulmonary venous pressures (Heath and Hicken, 1960). In severe mitral stenosis of many years' standing there are obvious structural changes in the pulmonary vessels but in acute left ventricular failure, it is conceivable that such a mechanism plays a part.

Distribution of Blood Flow in Saline-Filled Lung

If it is true that the uneven distribution of blood flow in the normal lung is caused by the effects of the hydrostatic pressure differences between the inside and outside of the blood vessels down the lung, there are only two situations where blood flow would be uniform. One is in the absence of a gravitational field as in outer space, and the other is if the airways were filled with a liquid which had the same specific gravity as blood, and the lung were suspended in a bath of this liquid. Under these conditions, the increasing hydrostatic pressure in the blood vessels down the lung would be balanced by a corresponding pressure outside the vessels.

Fig. 13 shows the distribution of blood flow in the isolated perfused dog lung which was inflated with saline and suspended in a saline bath. The measurements were made by injecting aggregated albumin labelled with ^{131}I (Taplin, Johnson, Dore and Kaplan, 1964) into the pulmonary artery and scanning the lung after the particles had been filtered out by the small pulmonary vessels. Subsequent scans were carried out after uniformly labelling the saline in the alveoli by rebreathing non-aggregated ^{131}I albumin. It can be seen that the distribution of blood flow was substantially uniform although there was an unexplained hump in the middle of the pattern; in any event there was no systematic difference down the lung. The vascular pressures used for these measurements were of the same order as those of Fig. 4A which shows the grossly non-uniform distribution of blood flow found in the air-filled lung under these conditions. Thus the measurements confirm the hypothesis that the uneven distribution of pulmonary blood flow is caused by hydrostatic effects (West, Dollery, Matthews and Zardini, 1966).

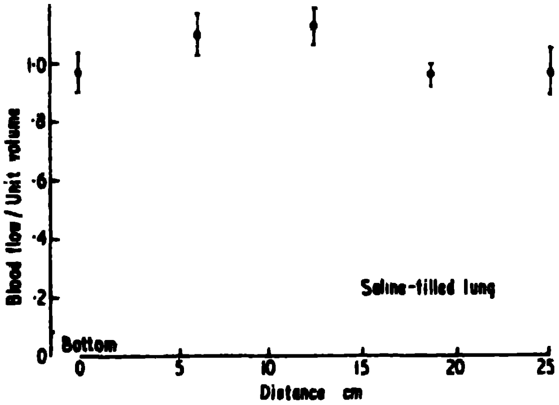


FIG. 13. Distribution of blood flow in isolated lung inflated with saline and suspended in a saline bath. Means and standard errors from eight lungs. Note that the distribution is substantially uniform and that there is no systematic change up the lung. (From *J. app. Physiol.*)

In summary, the uneven topographical distribution of blood flow in the lung is an inevitable consequence of the intimate apposition of blood and air which results in large pressure differences across the pulmonary vessels at different levels in the lung. The resulting imbalance of blood flow and ventilation causes large regional differences in gas exchange but interference with overall gas transfer is negligible unless the lung is exposed to an increased gravitational field, or pulmonary hypotension occurs through disease. The uneven distribution of blood flow results in peculiar overall pressure–blood flow relations in that the pulmonary vascular resistance falls greatly when either the pulmonary arterial or venous pressures are raised. These properties are advantageous to the lung, the one organ of the body which receives the whole cardiac output during a variety of physiological states. Interstitial oedema may invert the normal pattern of blood flow by increasing the vascular resistance in the dependent zone of the lung. The observation that blood flow is uniformly distributed in the saline-filled lung is added evidence that hydrostatic effects are responsible for the uneven distribution of blood flow in the air-filled lung.

ACKNOWLEDGEMENTS

The radioactive gas measurements were made in conjunction with Drs C. T. Dollery, P. Hugh-Jones, N. L. Jones, C. M. E. Matthews, A. Naimark and P. Zardini, and with the co-operation of the M.R.C. Cyclotron Unit. The work was supported by the Medical Research Council.

REFERENCES

BALL, W. C., STEWART, P. B., NEWSHAM, L. G. S. and BATES, D. V. (1962) *J. clin. Invest.*, **41**, 519.

BANISTER, J. and TORRANCE, R. W. (1960) *Quart. J. exp. Physiol.*, **45**, 352.

BARR, P. O. (1963) *Acta physiol. scand.*, **58**, suppl. 207.

BEVEGARD, S., HOLMGREN, A. and JONSSON, B. (1960) *Acta physiol. scand.*, **49**, 279.

BITTER, H. S. and RAHN, H. (1956) *Studies in respiratory physiology.* Wright Air Develop. Ctr. 3rd series. 1, Dayton: Ohio.

BRYAN, A. C., BENTIVOGLIO, L. G., BEEREL, F., MACLEISH, H., ZIDULKA, A. and BATES, D. V. (1964) *J. appl. Physiol.*, **19**, 395.

DOLLERY, C. T. and WEST, J. B. (1960) *Circ. Res.*, **8**, 765.

FREEMAN, J. and NUNN, J. F. (1963) *Clin. Sci.*, **24**, 135.

GERST, P. H., RATTENBORG, C. and HOLADAY, D. A. (1959) *J. clin. Invest.*, **38**, 524.

HEATH, D. and HICKEN, P. (1960) *Thorax*, **15**, 54.

HINSHAW, L. B., DAY, S. B. and CARLSON, C. H. (1959) *Amer. J. Physiol.*, **197**, 309.

HOWELL, J. B. L., PERMUTT, S., PROCTOR, D. F. and RILEY, R. L. (1961) *J. appl. Physiol.*, **16**, 71.

MACKLIN, C. C. (1946) *Rev. Can. Biol.*, **5**, 199.

ORTH, J. (1887) *Ätiologisches und Anatomisches über Lungenschwindsucht.* A. Hirschwald, Berlin, page 20.

PERMUTT, S., BROMBERGER-BARNEA, B. and BANE, H. N. (1962) *Med. Thorac.*, **19**, 239.

STAUB, N. C. and STOREY, W. F. (1962) *J. appl. Physiol.*, **17**, 381.

TAPLIN, G. V., JOHNSON, D. E., DORE, E. K. and KAPLAN, H. S. (1964) *Report of I.A.E.A. Symposium on Medical Radioisotope Scanning.* Athens.

TOCKER, A. M. and LANGSTON, H. T. (1952) *J. thorac. Surg.*, **23**, 539.

TROWELL, O. A. (1943) *Quart. J. exp. Physiol.*, **32**, 203.

WEST, J. B. (1962) *J. appl. Physiol.*, **17**, 893.

WEST, J. B. and DOLLERY, C. T. (1960) *J. appl. Physiol.*, **15**, 405.

WEST, J. B. and DOLLERY, C. T. (1965) *J. appl. Physiol.*, **20**, 175.

WEST, J. B., DOLLERY, C. T. and HEARD, B. E. (1965) *Circ. Res.*, **17**, 191.

WEST, J. B., DOLLERY, C. T. and NAIMARK, A. (1964) *J. appl. Physiol.*, **19**, 713.

WEST, J. B., DOLLERY, C. T., MATTHEWS, C. M. E. and ZARDINI, P. (1966) *J. appl. Physiol.* In press.

WEST, J. B. and JONES, (N. L. 1965) *J. appl. Physiol.* **20**, 825.

WOOD, E. H., NOLAN, A. C. and DONALD, D. E. (1963) *Aerospace Med.*, **34**, 270.

XVIII

Tissue Oxygen Requirements

I. S. LONGMUIR

THE oxygen requirements of animals can be considered in terms of the ability of the pulmonary and cardiovascular systems to supply oxygen to the tissues and the tissue oxygen requirements. This last factor can be defined as the amount of oxygen needed in unit time and the tension at which it must be supplied. The problem of tissue oxygen requirements can be approached in two ways. Isolated enzyme systems or pieces of tissue can be studied.

Many of the enzymes which react with oxygen have been extracted from tissues and their activity and affinity for oxygen have been determined *in vitro*. In order to use this information to arrive at values for the oxygen requirements of whole tissues it is necessary to know the distribution of these enzymes within tissue and the path of oxygen transport between them and the capillary blood. Thus both intact tissues and enzymes must be studied in order to understand tissue oxygen requirements.

Two experimental techniques have been developed to measure oxygen uptake by biological systems, manometry and polarography. Of these the latter has proved of great value in measuring the parameters of tissue oxygen requirements.

The first type of polarographic system used to study respiration was one employing the dropping mercury electrode. However this system has some disadvantages and it has now been replaced by systems using the solid electrode. It is necessary to understand something of the basic principles involved in the operation of this type of instrument since the technique may influence the biological system and vice versa.

When a solid electrode is immersed in a solution of oxygen and made sufficiently negative with respect to the solution, oxygen is reduced according to the following equations:

$$O_2 + e = O_2^- \qquad (1)$$

$$O_2^- + e = O_2^{--} \qquad (2)$$

$$O_2^{--} + 2H^+ = H_2O_2 \qquad (3)$$

$$O_2^- + H_2O_2 = O_2 + {}^-OH + {}^\cdot OH \qquad (4)$$

$${}^\cdot OH + e = {}^-OH \qquad (5)$$

It will be seen that if all of the hydrogen peroxide formed in equation (3) undergoes reaction according to equation (4) then water will be formed by the reduction of oxygen and four electrons will be required to reduce each molecule. In practice some of the hydrogen peroxide escapes and so the number of electrons involved in the reduction of each oxygen molecule is less than four. Using gold as the electrode it is about 3·7. It is at this point that one of the chief sources of error in the use of a solid polarographic electrode can occur. If the biological system contains hydrogen peroxide or catalase it may affect the polarographic signal or the hydrogen peroxide formed by the electrode may affect the biological system.

In order to use the reduction of oxygen at the gold electrode to measure the amount of oxygen in solution it is necessary to consider one other principle. When a solid surface is moved rapidly in relation to a liquid the bulk of the liquid is made turbulent, but a thin layer of virtually stationary fluid surrounds the solid surface. If now a potential difference is applied between the gold and the liquid sufficient to reduce all the oxygen at the interface more oxygen will diffuse from the bulk of the liquid across the stationary layer to the gold where it will be reduced. When this diffusion gradient is established oxygen will be arriving at the electrode surface at a rate proportional to its concentration in the bulk of the solution and the current flowing across the electrode will be proportional to the rate of reduction of oxygen at the surface. Thus the following relation between current flowing and oxygen content can be formulated:

$$i_d = pO_2ADnF/\delta \qquad (6)$$

where i_d = polarographic diffusion current
A = area of the electrode
D = diffusion coefficient in Krogh's units i.e. $cm^3gas/cm^2/minute/atmosphere$ partial pressure/cm
pO_2 = partial pressure of oxygen
n = number of electrons involved in reduction of one molecule of oxygen
F = Faraday (96,500 Coulombs)
δ = thickness of stationary layer of a slice

In practice it is not possible to measure all the constants accurately. Experimental measurements of i_d and pO_2 show that there is a linear relationship between them from 0·01–2,000 mm Hg.

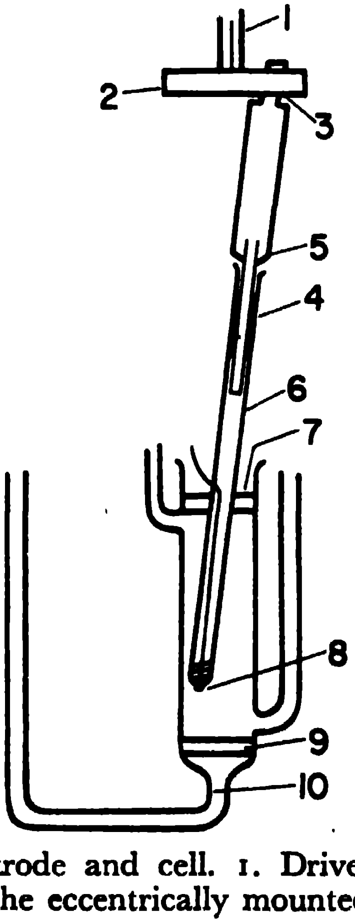


FIG. 1. Rotating electrode and cell. 1. Drive shaft of a synchronous motor. 2. Disc housing the eccentrically mounted self-centring bearing 3. The drive rod 5 is connected to the glass tube 6 through a rubber mounting 4. The tube is inserted into the cell through a rubber disc 7 and the gold electrode 8 is fused into this tube. 9 is a sintered glass plate separating the contents of the cell from the KCl-agar bridge 10 which connects to the reference anode.

In Fig. 1 can be seen one type of electrode system (Longmuir, 1957). With suitable precautions this assembly gives reproducible polarograms and it is possible to obtain diffusion currents which vary by less than 1 per cent in a period of days.

Since the volume of the cell is fixed and it is full of liquid with no gas phase the rate of fall of partial pressure will be proportional to the rate of removal of oxygen from the cell. This electrode responds very rapidly to changes in oxygen tension. This is a property of a bare rotated polarographic electrode of this type. Stationary bare electrodes do not have such good response times and covered electrodes are even slower, some have 95 per cent response times of the order of seconds.

Thus by measuring the rate of change of current with this device it is possible to measure the respiration rate of a biological system suspended in the polarographic cell. At the same time the partial pressure of oxygen being known it is possible to construct a respiration rate versus partial pressure of oxygen curve for that particular biological system.

In this way the respiration rate of a number of systems as a function of partial pressure of oxygen have been measured, Figs. 2, 3. They all give the same general picture. At zero oxygen tension the respiration rate becomes zero but with increasing tension the rate rises rapidly at first and then more slowly to a maximum rate at the critical oxygen tension, that is the lowest tension at which the system respires at its maximum rate. Although all the systems studied give the same shaped relationship it is possible to discern important differences. The outstanding difference lies in the scale. With some systems the maximum rate is achieved at very low tensions while with others it is a thousandfold higher. There are also significant differences in the shape of the curve below the maximum rate.

In order to try to explain these differences in terms of diffusion and chemical kinetics the respiration rates as functions of partial pressure of oxygen of a number of systems at various levels of organization have been studied. The systems are: electron transport particles, heart muscle preparation, mitochondria, individual cells and tissue slices. The first two represent essentially cytochrome oxidase plus the cytochrome system. These preparations appear to be respiring systems in

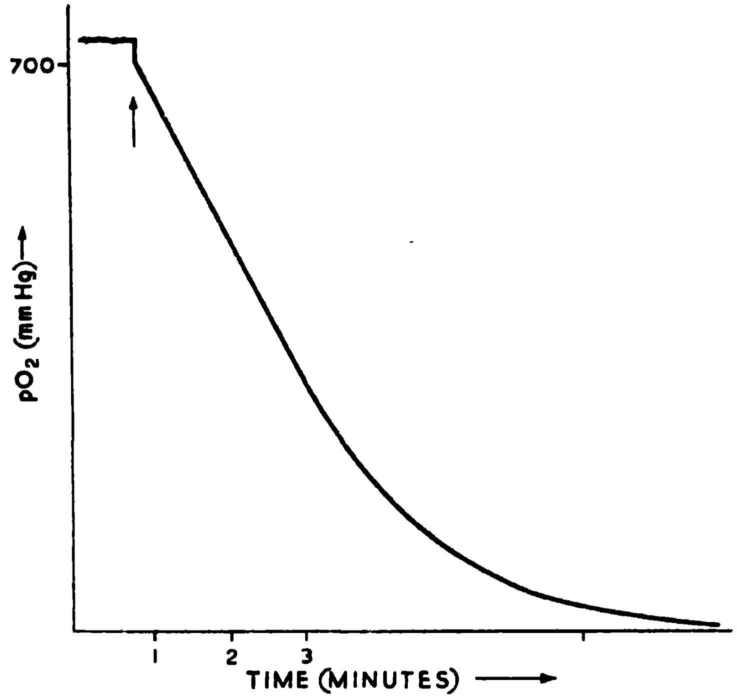


FIG. 2. A record of the diffusion current against time. Initially the line is horizontal when the cell contains only oxygenated saline. At the arrow the biological system is inserted into the cell. The tension then falls at a diminishing rate till no oxygen remains.

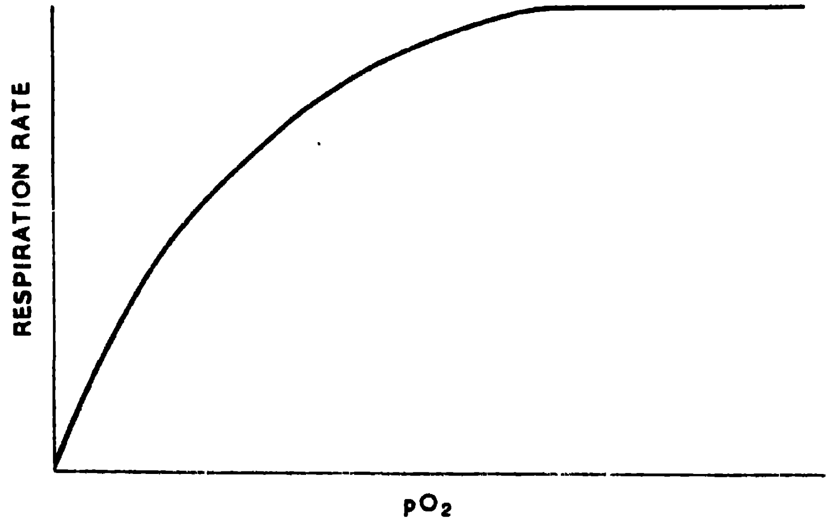


FIG. 3. This curve is obtained from the previous one by plotting the slope of the trace which is proportional to the respiration rate as a function of partial pressure of oxygen.

which diffusion plays the smallest part. In the case of the next, mitochondria, there appears to be the possibility of a gradient of oxygen tension between the outside of the mitochondria and the cytochrome oxidase situated within it. That is to say that now diffusion may play a part in determining the oxygen requirements of this system. In the case of the cells and slices it would be expected that diffusion would play an increasingly important role.

The results of a typical experiment are shown in Fig. 2. If we take from the curve the partial pressure of oxygen at which the system is respiring at half its maximum rate and compare the values obtained from the different systems it should be possible to obtain some measure of the gradients of oxygen tension through a tissue. Some experimental results obtained with this technique are shown in Table 1. There is a substantial rise in the half rate partial pressure of oxygen between electron transport particles and heart muscle preparation on the one hand and mitochondria on the other 0·03–3·0 mm. This can be interpreted as a measure of the partial pressure gradient between the outside of mitochondria and the oxidase site within. However between mitochondria and individual cells there appears to be no gradient. The difference between the value for cells and slices (3–100 mm) suggests a significant gradient here.

TABLE 1

System	Km (mmO_2)
Electron Transport Particles	0·03
Heart Muscle Preparation	0·03
Liver Mitochondria	2·8
Liver Cells	3·0
Liver Slices	100·0

It would appear from these results that there is a significant gradient of oxygen tension through tissue from the outside to the site of oxygen consumption within the cell. However when these figures plus the respiration rate of the tissue and its dimensions are used to calculate the diffusion coefficient of oxygen through tissue figures are obtained, which vary by one hundredfold. Closer examination of the results yields another discrepancy. Warburg (1930) has developed an equation to calculate the oxygen tension within a slice. This equation can be rearranged to give the critical oxygen tension of a slice as a function of slice thickness, unit respiration rate, and the diffusion coefficient of oxygen. It is:

$$C_s = \frac{\alpha H^2}{8D} \tag{7}$$

where α is the respiration rate in mls O_2/minute/ml of tissue
H is the thickness in cms
D is the diffusion coefficient in mls/cm^2/minute/atmosphere/cm

That is to say that the critical oxygen tension of a slice, the lowest tension at which it will just respire at its maximum rate, should be proportional to the square of its thickness. However a number of determinations of the critical oxygen tension obtained with slices of thickness varying from 0·1–0·5 mm shows at most only a threefold variation even from different tissues.

The critical oxygen tension appears to be largely independent of slice thickness. We thus have a number of paradoxes.

(1) There appears to be no gradient of oxygen tension across individual cells.

(2) There appears to be a very steep gradient across slices.

(3) A substantial increase in thickness of the slice does not increase the total fall in oxygen tension by the predicted amount, thus the gradient appears to fall in proportion to the distance through the tissue.

This straightforward approach to the problem of the diffusion of oxygen in the tissues cannot be valid. It may be that the assumption that cytochrome oxidase is the only important oxygen acceptor in tissue is not valid. That is to say that in going from the electron transport particles up to tissue slices we were not only adding geometrical complexity but something else as well. There might be in tissue a relatively large amount of an oxygen acceptor with such a low affinity for oxygen that this limited the respiration rate and the diffusion processes were of secondary importance. If such an enzyme system does exist then the critical oxygen tension of a slice of any thickness as calculated from Warburg's equation would be raised by a substantial and constant amount so that the increase of critical oxygen tension on increasing the slice thickness might become negligible.

This possibility can be investigated by applying Warburg's equation to the problem in a different way. Before, it was assumed that the critical oxygen tension of the oxidase system is negligible less than 1 mm. If we do not make this assumption then the value of the critical oxygen tension of a slice from Warburg's equation becomes:

$$C_s = \frac{\alpha H^2}{8D} + C_c \tag{8}$$

where C_c is the critical oxygen tension of the terminal oxidase and the other symbols have the same meaning as before.

There are two unknowns in the equation so a single determination will not be enough to evaluate them. The unknowns can be determined by plotting the square of the slice thickness against the critical oxygen tension divided by the respiration rate of unit volume for a large number of determinations. This has been done for liver, kidney and heart slices and the results are shown in Table 2.

TABLE 2. Diffusion coefficients of oxygen through the three tissues and the critical oxygen concentration of the terminal oxidase calculated from the slopes and intercepts of the regression lines (d^2 on C_s/Q_{O_2})

Organ	10^{-4} × slope	10^{-4} × S.E. of slope	$10^4 D_{O_2}$ (Krogh's units)	Intercept	S.E. of intercept	C_c(mM)
Kidney	1·52	0·29	0·63	-0·048	0·120	0·093
Liver	3·92	0·88	1·64	-0·140	0·141	0·168
Heart	3·74	0·74	1·54	-0·093	0·108	0·126

It will be seen that the diffusion coefficient now appears to be nearly ten times higher than that for water and there does appear to be an oxidase in tissue which has rather a low affinity for oxygen. $C_c = 400 - 200$ mm.

The high figure for the diffusion coefficient can be explained if there is an appreciable amount of stirring within the cell. It is likely that this is the case since mitochondria appear to imbibe water continuously while actively extruding it. This possibility was investigated by attempts to increase the viscosity of protoplasm by the method of Seifriz (1941). He showed that anaesthetic agents would slow or stop the protoplasmic streaming of slime moulds so the effect of carbon dioxide, ether and chloroform on the critical oxygen tension of tissue slices was investigated. All these agents were found to increase the critical oxygen tensions by amounts which would correspond to the reduction of the diffusion coefficient of oxygen to near that for water (Longmuir and Bourke, 1960; Longmuir, 1961; McCabe, 1964).

It has long been assumed that practically all tissue oxygen consumption goes through cytochrome oxidase so it seems improbable that there could be an oxidase with a low affinity for oxygen in sufficient amount in the tissue to have this strong effect on its critical oxygen tension. However, it has recently been shown by Longmuir and McCabe (1964) that the proportion of tissue oxygen consumption going through cytochrome oxidase in liver cannot exceed 44 per cent and may be about 25 per cent.

These two findings that the diffusion coefficient of oxygen in tissue is high and that there is at least one oxidase system in tissue which requires a high partial pressure of oxygen for optimal activity raises the question as to how important the diffusion of oxygen in tissue is in limiting respiration rate in oxygen lack.

The ideal approach would be one in which all the relevant parameters, the affinity of the oxidases and oxygenases, the geometry of the tissue and the diffusion coefficients were given due weight. An equation has been developed by O'Sullivan which relates all these factors. Unfortunately this equation is insoluble.

Although both diffusion and kinetics must play some part in limiting respiration rate a system with an oxidase having a very low affinity for oxygen and a high diffusion coefficient the kinetic processes may be so dominant that effectively they limit respiration.

The problem has been investigated by examining the consequences of either possibility (Longmuir and McCabe, 1965). If theoretical relationships could be set up for model systems in which either diffusion or kinetics were limiting and these relationships were sufficiently divergent to be susceptible of experimental investigation then the problem could be solved.

An attempt to find a theoretical relation between respiration rate and partial pressure of oxygen has been made by Rashevsky (1960) for spherical organisms assuming only diffusion as being rate limiting. He assumed that there would be an appreciable barrier to diffusion at a membrane surrounding the sphere while within there would be a uniform diffusion coefficient. However, it now appears that the cell contains many membranes through which oxygen must diffuse so it seems reasonable to take a mean diffusion coefficient. When this is done Rashevsky's rather cumbersome equation reduces to:

$$v = \frac{\bar{V}pO_2}{k + pO_2} \tag{9}$$

where v is the respiration rate at the given pO_2

$\bar{V}$ is the maximum respiration rate
k is a function of the diffusion coefficient of oxygen through the tissue.

This has exactly the form of the Michaelis-Menton equation.

All experimental work on the relationship between partial pressure of oxygen and the activity of respiratory enzymes shows that it has the form of Michaelis-Menton kinetics (Laser, 1952; Longmuir, 1954; Fridovich and Handler, 1962) that is to say that the rate xpO_2 curve has the general shape shown in Fig. 4A.

Thus with a spherical model it is impossible to distinguish between kinetic and diffusion limitation of respiration rate. Clearly if enzyme kinetics limit respiration the relation between rate and pO_2 will have this same form whatever the geometry of the model.

In the case of the spherical model the diffusion paths are convergent. It seemed reasonable therefore to consider models having parallel and divergent paths.

Tissue model with parallel diffusion of oxygen

The simplest model with such a path is a tissue slice of uniform thickness and large surface area in relation to the thickness.

Warburg (1930) showed that oxygen would diffuse into such a slice to a depth of

$$\sqrt{\frac{2DpO_2}{\alpha}}$$

where pO_2 is the partial pressure of oxygen in contact with the slice

D is the diffusion coefficient of oxygen

α is the respiration rate for unit volume of tissue.

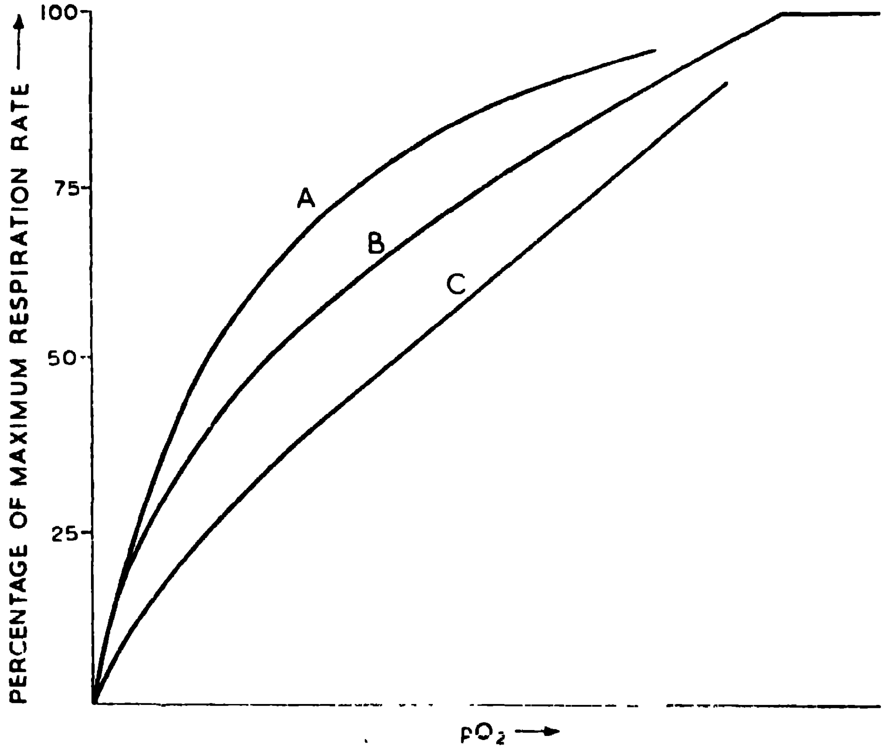


Fig. 4. Relation of respiration rate of oxygen tension. A. when enzyme kinetics are limiting, or diffusion into a sphere; B. when diffusion of oxygen into a slice is limiting; C. when diffusion of oxygen out of a cylinder is limiting.

Thus for a slice of thickness H and area A suspended in a stirred solution the respiration rate would be:

$$v = 2\alpha A\sqrt{\frac{2DpO_2}{\alpha}} \tag{10}$$

Thus the respiration rate of a slice should rise as the square root of the partial pressure of oxygen. With rising oxygen tension, oxygen would penetrate further into the slice until the fronts from each side meet. Now all the slice would contain oxygen and if enzyme kinetics were not limiting the maximum rate would be achieved. Thus the critical oxygen tension of a slice would be given by:

$$\frac{H}{2} = \sqrt{\frac{2Dp^1O_2}{\alpha}} \qquad (11)$$

$$p^1O_2 = \frac{H^2\alpha}{8D} \qquad (12)$$

The form of the relation given by equation (10) is shown in Fig. 4B.

Tissue model with divergent diffusion of oxygen

Although there are several models which fulfil this criterion, that of diffusion of oxygen out of a capillary has the most immediate relevance to the current problem.

Krogh (1919) has developed an equation relating the distance R to which oxygen will diffuse into respiring tissue from a capillary of radius r:

$$pO_2 = \frac{\alpha}{D}\left(1{\cdot}15R^2\log\frac{R}{r} - \frac{R^2 - r^2}{4}\right) \qquad (13)$$

Using the same general procedure as before it is possible from this to construct a respiration rate against a pO_2 curve for this model. Fig. 4C. However, it is not possible to construct the complete curve in this case. When the advancing fronts of oxygen from adjacent capillaries meet the form will change. If the capillaries are uniformly distributed this will occur at 89·5 per cent of the maximum rate and presumably the rest of the curve will become asymptotic to the maximum.

EXAMINATION OF EXPERIMENTAL RESULTS

It might seem that the model system with divergent diffusion would be the best to examine since curves A and C differ most. However, in the case of the capillary there are further parameters to consider. For example, the fall in oxygen tension down the capillary which, since the oxyhaemoglobin dissociation curve is sigmoid, will not be linear. The problems associated with this approach have been discussed at length by

Brauer (1963). It seems that till more information is available no accurate measurements will be possible, so for the moment we are restricted to a consideration of a slice.

In this case diffusion limited respiration would result in a reasonably well defined critical oxygen tension; the lowest tension at which the maximum respiration rate is still obtained. This value should rise as the square of the slice thickness. If the rate is kinetically limited a critical oxygen tension would be hard to discern and virtually independent of slice thickness. Below this critical tension the curves would have a different form but this might not be apparent from inspection since it is not easy to compare curves. However curves A and B can be converted to linear plots. Lineweaver and Burk (1934) showed that Michaelis kinetics give a straight line if the reciprocal of rate is plotted against the reciprocal of substrate concentration, in this case pO_2. It can be seen from equation (10) that curve B would give a straight line of slope one half if log rate is plotted against log pO_2. Thus by determining which method gives a straight line when applied to experimental determinations of tissue slice respiration rate as a function of oxygen tension it should be possible to determine whether the respiration is diffusion or kinetically limited.

A number of plots of the respiration rate of slices of various tissues as a function of partial pressure of oxygen were obtained using the polarographic technique. Various slice thicknesses were used.

The results were replotted according to the method of Lineweaver and Burk (1934) and Longmuir and McCabe (1965). The reciprocal plots all gave straight lines while the log plots gave curves Figs. 5 and 6. Thus it appears that tissue respiration rate is limited in oxygen lack by the affinity of a terminal oxidase for oxygen rather than by diffusion gradients of oxygen through the tissue. In order to try to exclude the possibility of an unrecognized artefact artificial slices were prepared. These were made by imbedding heart muscle preparation in stiff agar. Thus they would correspond to tissues having a terminal oxidase with a high affinity for oxygen and having themselves a rather low diffusion coefficient. When the respiration rate of these slices as a function of oxygen tension was determined and the

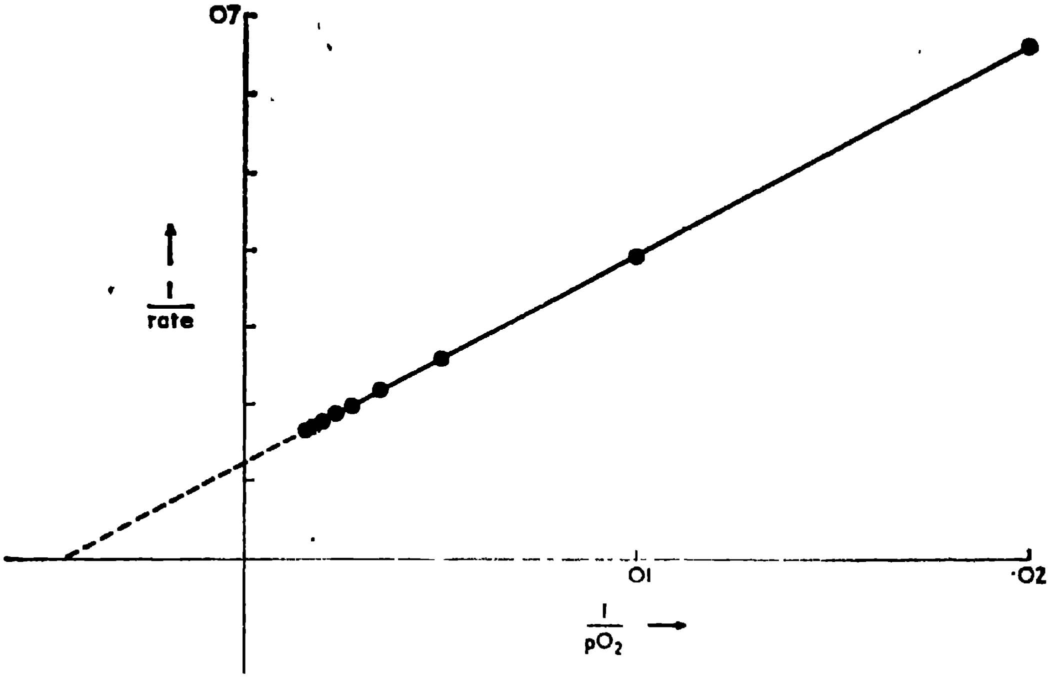


FIG. 5. Lineweaver-Burk plot for liver slice respiration

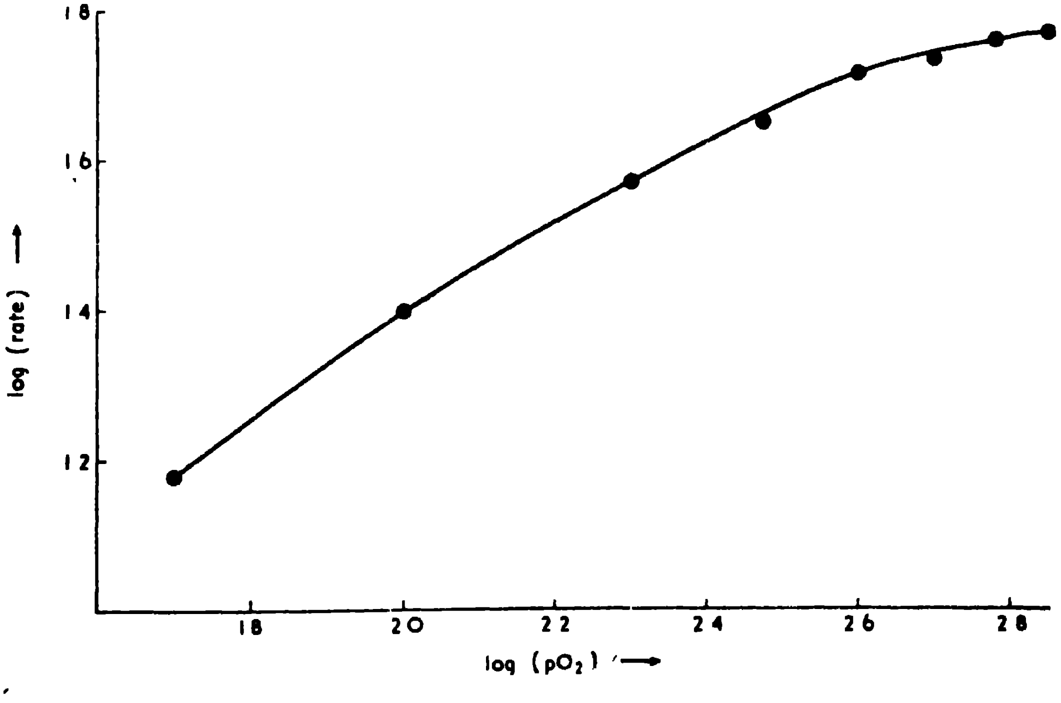


FIG. 6. Log plot for liver slice respiration.

results plotted on the reciprocal and log methods a completely different picture emerged Figs. 7 and 8. The former was curved and the latter straight. At the same time these slices had critical oxygen tensions which rose as the square of the slice thickness.

The fact that artificial slices behave in the predicted manner when examined in this way is of particular interest. Until recently, all theoretical attempts to relate respiration rate to partial pressure of oxygen have assumed that tissues conform to the artificial tissue we have prepared. Up to a point, therefore, our findings with this artificial system is a vindication of these theoretical treatments. Unfortunately, these models bear little relation to reality.

The finding that the respiration rate of real tissues conforms to Michaelis kinetics raises a serious difficulty however. More than thirty oxidases and oxygenases (the enzymes which react with molecular oxygen) have been described in tissue. Although the activities of each of these enzymes conform to Michaelis kinetics they have widely varying affinities and the sum of their activities should not approximate to Michaelis kinetics. In addition, there is the observation of Longmuir and McCabe (1964) that the proportion of oxygen going through cytochrome oxidase, which has the highest affinity for oxygen varies from 44 per cent–95 per cent with the age of the slice, yet this does not manifest itself in the expected change in kinetics.

This anomaly can be resolved by postulating an oxygen carrier, which behaves kinetically in the same way as an enzyme, between the external source of oxygen and all of the terminal oxidases and oxygenases in the tissue. This carrier would have to transport virtually all the oxygen, little could move by passive diffusion otherwise the experimental results would lie between the diffusion and kinetically limited picture.

The nature of such a carrier remains obscure. Scholander (1960) has shown that haemoglobin and myoglobin could act in this way but not in the conditions obtaining in the experiments described in this paper (Hemmingsen and Scholander, 1960). However, such a hypothesis could explain the findings of Longmuir and Bourke (1960) without the need for two hypotheses, namely a high diffusion coefficient and a low affinity oxidase through which the bulk of the oxygen passed.

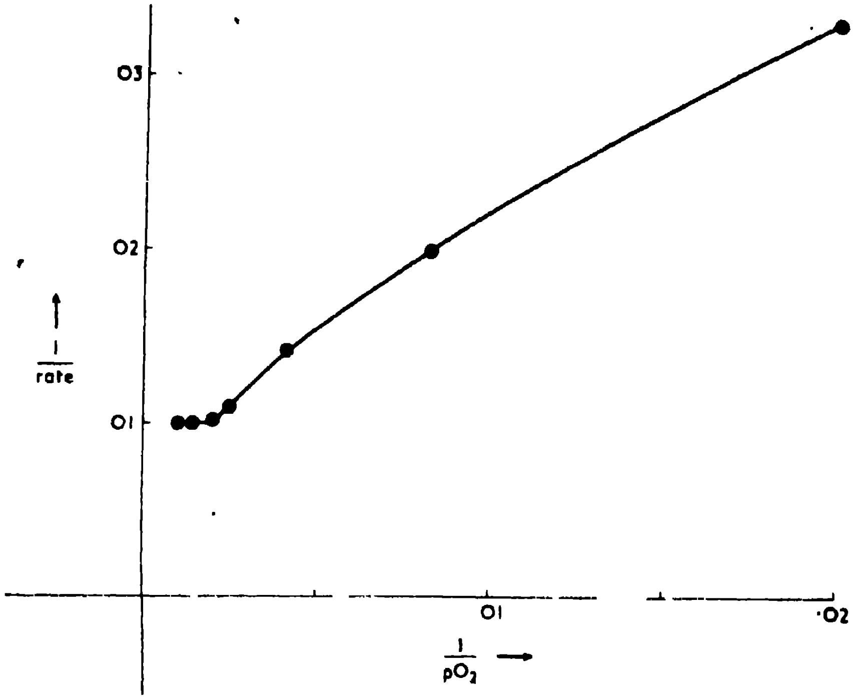


FIG. 7. Lineweaver-Burk plot for artificial slices.

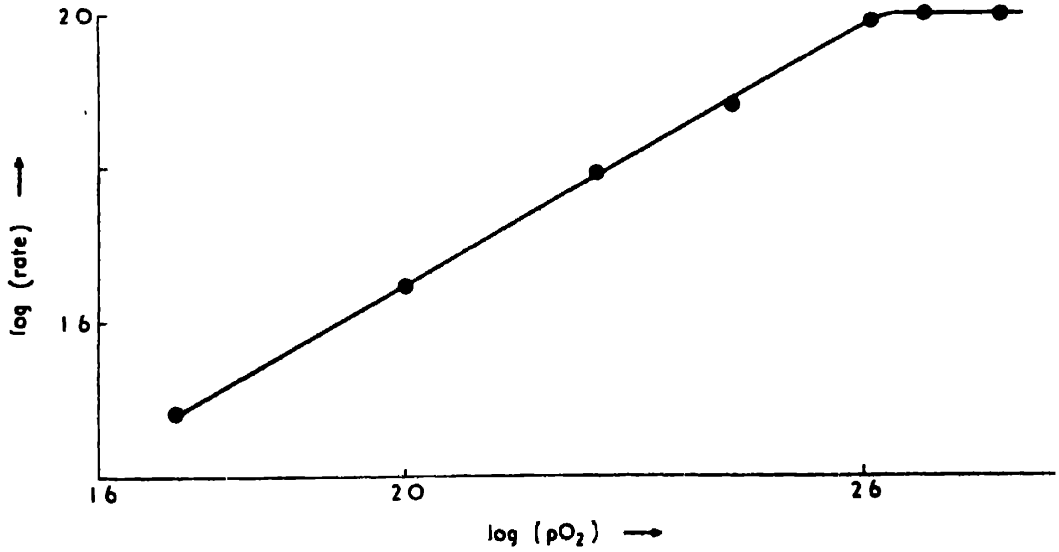


FIG. 8. Log plot for artificial slices.

The affinity of the oxygen carrier for oxygen can be determined from the Lineweaver and Burk plot. Values for different tissues lie in the range 80–370 mm. Thus the carrier cannot be either haemoglobin or myoglobin.

The problem of tissue oxygen requirements is clearly far from solved. Although we now have a great deal of detailed knowledge on the activity of the various enzymes which react with oxygen it has not yet proved possible to relate this precisely to events in intact tissue. When we try to do this a number of paradoxes appear. At the moment it seems that the only way to resolve some of them is by postulating a carrier system. Further work is clearly required in this area.

ACKNOWLEDGEMENTS

Fig. 1 appeared in the *Biochemical Journal*, Fig. 4 in the *Journal of Theoretical Biology* and Figs. 5–8 in the *Journal of the Polarographic Society*, and I wish to thank the Editors for their permission to reproduce them here.

REFERENCES

BRAUER, R. (1963) *Physiol. Rev.*, **43**, 115.
FRIDOVICH, I. and HANDLER, P. (1962) *J. biol. Chem.*, **237**, 916.
HEMMINGSEN, E. and SCHOLANDER, P. F. (1960) *Science*, **132**, 1379.
KROGH, A. (1919) *J. Physiol.*, **52**, 391.
LASER, H. (1952) *Proc. Roy. Soc. Ser. B.*, **140**, 230.
LINEWEAVER, H. and BURK, D. (1934) *J. Amer. Chem. Soc.*, **56**, 658.
LONGMUIR, I. S. (1954) *Biochem. J.*, **57**, 81.
LONGMUIR, I. S. (1957) *Biochem. J.*, **65**, 378.
LONGMUIR, I. S. (1961) 'Effect of Anaesthetic Agents on Tissue Respiration'. *Proc. Vth Int. Congress.* Moscow.
LONGMUIR, I. S. and BOURKE, A. (1960) *Biochem. J.*, **76**, 225.
LONGMUIR, I. S. and MCCABE, M. (1964) 'Cyanide and Azide Insensitive Tissue Respiration'. *VIth Int. Cong. Biochem.* New York. X-43.
LONGMUIR, I. S. and MCCABE, M. (1965) *J. theoret. Biol.*, **8**, 124.
MCCABE, M. (1964) Ph.D. thesis, University of London.
RASHEVSKY, N. (1960) *Mathematical Biophysics: Physico-Mathematical Foundations of Biology*, vol. i, p. 29. Dover Publications, New York.
SCHOLANDER, P. F. (1960) *Science*, **131**, 585.
SEIFRIZ, W. (1941) *Anaesthesia*, **2**, 300.
WARBURG, O. (1930) *Metabolism of Tumours*. Springer, Berlin.

XIX

Thyroid and Central Nervous Development

J. T. EAYRS

It has long been recognized that a close relationship obtains between the activity of the thyroid gland and cerebral function. Thyroid dysfunction in man gives rise to a well documented symptomatology (see, for example, Means, 1948; Werner, 1962) which, when the condition arises in the mature individual, can largely be attributed to metabolic factors and can be corrected by procedures designed to adjust the level of circulating thyroid hormone to meet the current needs of the tissues. This explanation is inadequate, however, to account for the effects of thyroid deficiency originating early in life when, in addition to an impairment to bodily growth and development, there arises a mental retardation whose severity is related to the age of onset of the disorder and which frequently does not respond to therapy. This difference of effect has given rise to the view that, when the developing brain is deprived of the influence of thyroid secretion, an irreversible abnormality of growth is superimposed upon the metabolic disorder found in the adult.

Such an interpretation has not always found general acceptance among clinicians for, although there is little scope for disagreement concerning the association of mental retardation with thyroid deficiency, there is a considerable literature which throws doubt on the importance of the time factor in the institution of therapy and hence on whether the cerebral defects can be attributed to the direct influence of thyroid deficiency during a critical period of development. The careful and detailed studies of Lewis (1937*a*,*b*), for example, failed to show any positive correlation between age of diagnosis, efficacy of treatment and restitution of normal mental development and similar findings have been reported from other sources (e.g. Brown, Bronstein and Kraines, 1939; Bruch and McCune, 1944; Radwin, Michelson, Kramer and Berman, 1949). Likewise, although the most recent and comprehensive study along these lines (Smith, Blizzard and Wilkins, 1957) showed a high correlation between the parameters of diagnosis and medication and the subsequent IQ attained by the patient, many apparently adequately treated cretins failed to respond.

The reasons for such apparent discrepancies have never been adequately resolved. It is possible that faulty initial diagnosis and the well-known propensity on the part of outpatients to conceal their failure to maintain prescribed treatment may be a contributory factor. Again, in instances where thyroid deficiency originates during the course of gestation, not only will the time of onset in relation to the state of cerebral growth generally be unknown but there is also uncertainty as to whether maternal hormone can cross the placental barrier in titres large enough to affect foetal development (Osorio and Myant, 1960). In this connexion, a possible basis for differentiating cretins which later respond to medication from those which do not is provided by the suggestion (McGirr and Hutchison, 1955) that a low level of foetal thyroid activity, though inadequate postnatally, might be sufficient to maintain cerebral maturation *in utero* whereas the completely athyreotic individual would be exposed to the effects of deficiency earlier in life. Yet another possibility is that the cretins failing to respond to treatment are those in which a primary amentia is genetically linked to the thyroid disorder (Tredgold, 1929) implying that any developmental changes attributable to the uncomplicated endocrine condition are, in themselves, reversible. There are, however, several grounds on which such an interpretation has been attacked. Thus there appears to be no greater than normal incidence of mental disability among the parents and siblings of cretins (Brown *et al.*, 1939); cases have been reported of cretinism arising in one only of a pair of monozygotic twins (Faxen, 1935; Pickering and Koulischer, 1956); sporadic cretinism is not associated with any increase in congenital abnormalities (Pickering and Fisher, 1958) and the response to treatment is similar to that seen in endemic cretinism (Jackson, 1934).

Any attempt to resolve these issues in terms of the anatomy and physiology of nervous tissues by recourse to clinical material is fraught with difficulty, for untreated cretinism is rare in sophisticated communities and few cretins come to autopsy during infancy. The problem is, however, one which might be expected to yield to controlled experimental analysis using laboratory animals and, beginning with a brief general assessment of the role of the thyroid during development, the following paragraphs will outline some of the attempts made along these lines to elucidate the factors involved.

THYROID AND MATURATION

Stemming from early work on the amphibian larva (Gudernatsch, 1912) it has become firmly established that interaction between thyroid hormone and some at present undefined but localized property inherent in the target tissues gives rise to a wide variety of physiological and anatomical changes all of which may be regarded as adaptations facilitating the transfer from an aquatic to a terrestrial mode of life, overtly manifest in the process of metamorphosis (Etkin, 1955). A similar role for the thyroid may be traced in mammalian forms for, although acting synergistically with other growth-promoting principles to facilitate the incremental phase of development (Evans, Simpson and Pencharz, 1939; Ray, Simpson, Li, Asling and Evans, 1950) the dominant and specific role of thyroid hormone is to regulate those changes in bodily form which mark the transition from infant to adult (Scow, Simpson, Asling, Li and Evans, 1949; Walker, Asling, Simpson, Li and Evans, 1952). Such maturational phenomena are particularly well illustrated by the effects of thyroid deprivation or excess during the early development of the rat as shown by Plates X–XI, Figs. 1 and 2.[1] In the former the short nose, closed eye, thickened partially-elevated pinna and fat appearance of the normal 10-day-old pup may be contrasted with the elongated snout, open eye, highly elevated pinna of adult proportions and lean adult-like bodily configuration of a littermate given 3 μg tri-iodothyronine daily from the day of birth. Plate XI, Fig. 2 serves, by contrast, to illustrate the reverse process in a pair of littermates one of which was thyroidectomized at birth.

---

[1] These plates will be found between pages 322–3.

In metamorphosing amphibia adaptations, comparable with those taking place at the periphery and which show a similar element of localization, are induced in the central nervous system. Some regions, notably the olfactory, visual and locomotor apparatus to all of which modification is necessary to equip the metamorphosed individual for survival in its new habitat (Tusques, 1955), appear to be selectively and directly sensitive to thyroid influence (Weiss and Rossetti, 1951) but it is by no means certain either that the thyroid is necessary for, or that its influence is confined to, phenomena involving structural reorganization. The accelerated appearance of reflex behaviour which follows the local application of thyroxine to the brain, for instance, follows a time course which suggests that the hormone may act, not so much as an organizer of neural mechanisms as by triggering into activity a pre-differentiated substrate (Kollros, 1943).

By analogy, it might be anticipated that the thyroid would exert upon the mammalian nervous system an influence, if not in detail at least in principle, comparable with that seen in lower forms. The rat, whose retarded physical development in the thyroid deficient state has already been noted, proves to be a most suitable subject for investigation of this possibility. It is readily reared in the laboratory producing litters of a size which permit of the use of littermate control techniques; its development is particularly susceptible to departures from the euthyroid state; it can without difficulty be thyroidectomized during infancy either by the chronic administration of goitrogen (Hughes, 1944), by giving a single injection of radio-iodine (Goldberg and Chaikoff, 1949) or by surgical removal of the gland. But perhaps the most important factor of all is that at the time of birth much of the central nervous system of the rat is poorly differentiated thus permitting the experimental investigation of the influence of thyroid deficiency (and, for that matter, excess) upon maturational events which, in man, occur during gestation. Against this background the species has been used to analyse the effects of departures from the euthyroid state upon cerebral development in terms of the structural correlates of changes in behaviour and the electrical activity of the brain; the extent to which such effects are mediated directly or indirectly and whether they may be regarded as specific attributes of the influence of the thyroid.

BEHAVIOUR

(a) *Innate*

The rate of central nervous maturation during the first days of life in the rat can be assessed in terms of the time of first appearance of a series of well defined landmarks such as the startle response to auditory stimuli and the capacity to land on all four feet when dropped back downwards and the elicitation of the placing reflex. These automatic responses emerge as the result of the maturation of complex sensory-motor integrative mechanisms which are absent in the newborn; all are grossly retarded by thyroidectomy and slightly advanced by giving thyroid hormone (Eayrs and Lishman, 1955; Eayrs, 1964). It is thus clear that the thyroid exerts a regulatory influence upon the development of nervous tissues but the fact that all these forms of innately organized behaviour do eventually appear in the repertory of the thyroidectomized individual suggests that this influence may be one of sensitizing or determining the rate of growth of a neural mechanism rather than of directing its patterning. Similar conclusions were reached by Hamburgh and Vicari (1957) following their studies with the mouse. These findings shed little light on the apparent irreversibility of mental retardation in the human cretins and greater interest must attach to the influence of the thyroid on behaviour which owes its origin to the history of the individual rather than to that of the species and in which the cerebral cortex plays a dominant role.

(b) *Adaptive*

By comparison with the effects of severe thyroid deficiency arising in the human adult those seen in the rat are slight and determined more by rate of performance than by ability (for literature, see Eayrs and Levine, 1963) a fact which enables impairment due to developmental causes to be seen more clearly. In spite of this, however, experiments designed to elucidate the importance of the thyroid secretion early in life on subsequent capacity for adaptive behaviour have at first sight yielded inconsistent results. In an early experiment, using an elevated T-maze set up in an environment presenting a wide variety of external cues, Eayrs and Lishman (1955) were able to demonstrate marked differences between the performance of rats thyroidectomized at birth and that of their normal littermates (Table 1 (A)). Although not taking significantly more trials to learn the problem to an acceptable criterion of success cretinoid rats made more errors in terms of time spent in the maze and took longer to learn, though losing less time over each error than the controls. The slower running of the hypothyroid animal has been observed by other workers in rats thyroidectomized during adult life and attributed to a reduced rate of metabolism. The increase in error scores, however, is peculiar to the rat thyroidectomized in infancy and would therefore seem to reflect the influence of thyroid deficiency during the period of cortical maturation. The permanence of the impairment is thrown open to question by the ability of medication with thyroxine, begun on the 25th day of age completely to restore the performance of the cretinoid rat to normal. This finding at first sight provides support for the view that deficiency of thyroid hormone *per se* is not the factor responsible for the mental retardation of cretinism and other, more recent, experiments (Eayrs and Levine, 1963) might be regarded as pointing in the same direction. In these the animals were required to avoid an electric shock in response to an auditory conditioned stimulus. Table 1 (B) shows the characteristics of the behaviour of neonatally thyroidectomized rats, unmedicated and medicated from the age of 5 months, compared with that of their normal littermates. The increased number of errors made by the experimental animals together with a reduced rate of performance and degree of alertness as measured respectively by latency of response and activity between trials, here parallel the maze-learning results not only in the nature of the deficiencies revealed but also in the complete eradication of these defi-

PLATE X

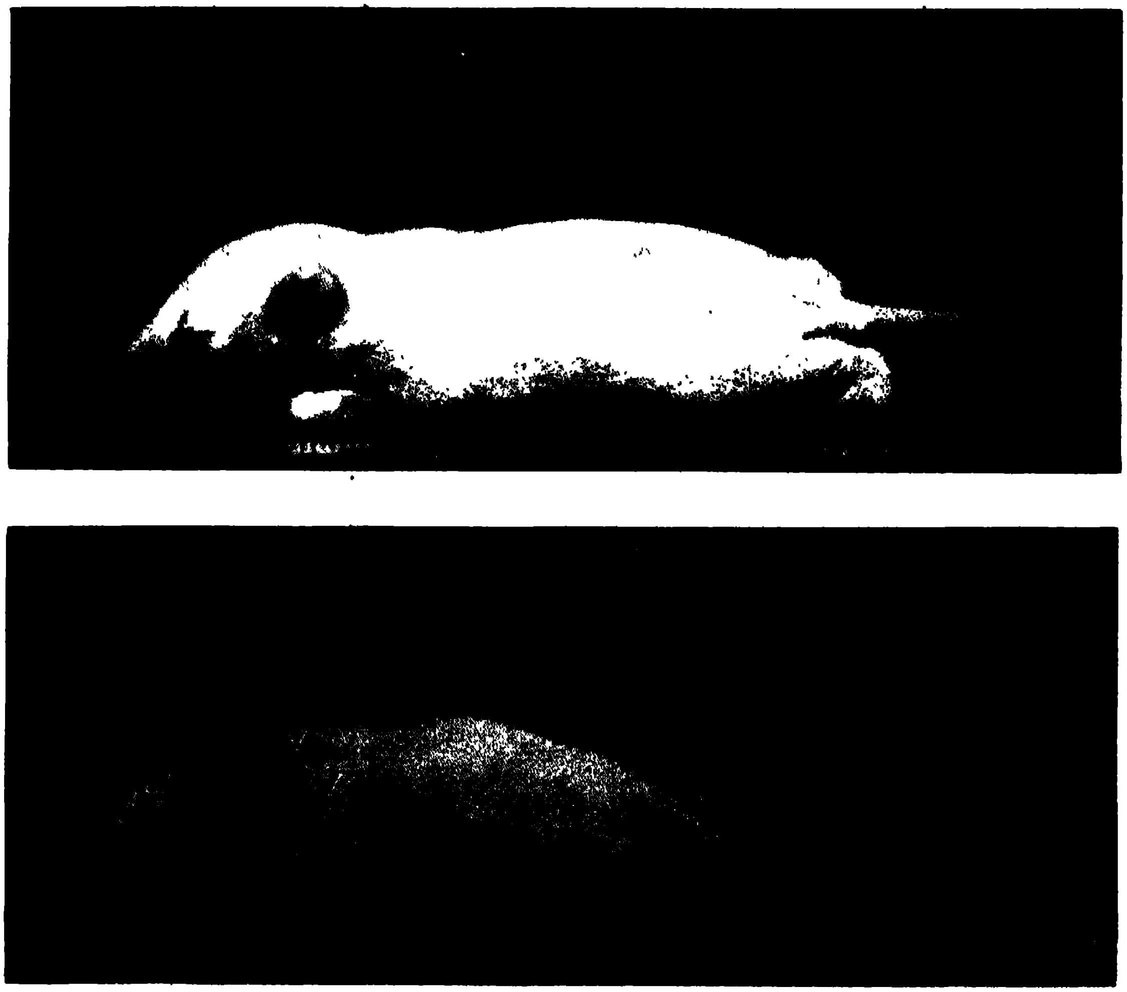


FIG. 1. Effect of giving tri-iodothyronine daily from birth on growth of rats. Littermates aged 10 days; treated rat below.

PLATE XI

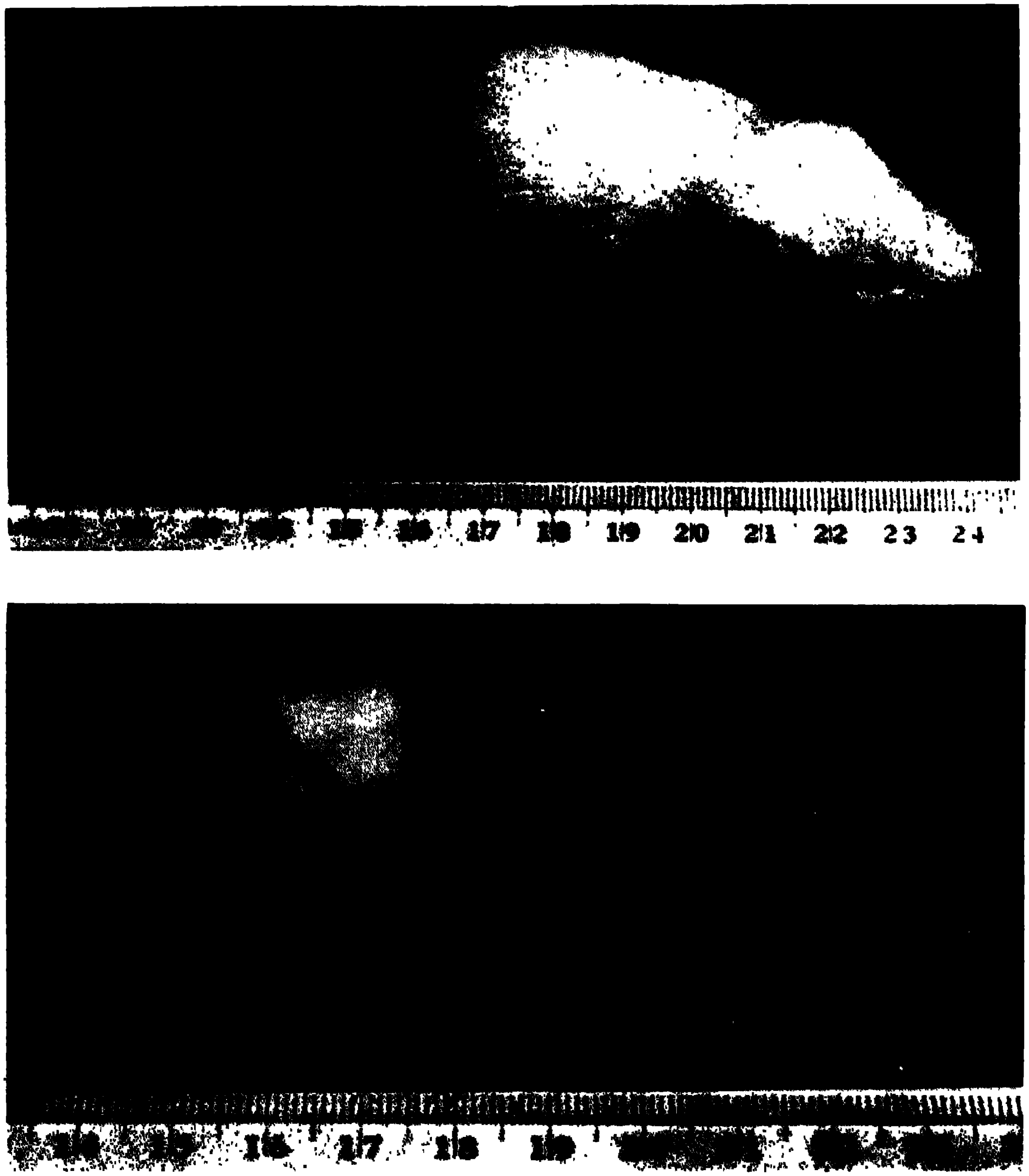


FIG. 2. Effect of thyroidectomy at birth on growth of rats. Littermates aged 25 days; thyroidectomised rat below.

TABLE 1. Effect of thyroid on learned behaviour

A. Elevated T-maze*

	Thyroidectomy at birth	Medicated from 25 days
Error scores	+11·3	- 5·1
Rate of performance:		
Time to learn	+10·7 mins	- 3·4 mins
Time for correct run	+27·1 secs	- 0·5 secs
Time lost per error	- 6·6 secs	- 10·6 secs

B. Conditioned avoidance†

	Thyroidectomy at birth	Medicated from 5 months
Error scores	- 32·6	- 0·2
Rate of performance:		
Time for 60 trials	+ 198·0 secs	+6·0 secs
Response latency—unconditioned	+ 2·0 secs	- 0·3 secs
Response latency—conditioned	+ 0·9 secs	0·0 secs

* Data of Eayrs and Lishman (1955).

† Data of Eayrs and Levine (1963).

All data represent mean differences in performance between experimental and control littermates. Effects of medication are in all instances statistically significant.

ciencies by medication started during adult life. Indeed, in this situation, the performance of rats thyroidectomized during infancy proved to differ only in severity from that of animals thyroidectomized when fully mature.

By contrast, however, using the Hebb-Williams closed-field test which has been successfully applied to the detection of small cortical lesions in the rat (Rabinovitch and Rosvold, 1951) and which may therefore be regarded as providing a more critical functional index, a very different pattern of results has emerged (Eayrs, 1961). The test consists of a series of twelve maze-like problems which are explored by the thirst-motivated rat in search for water. Each animal is given eight successive attempts at each problem and its performance measured in terms of the number of errors, or occasions on which it departs from the direct route from start to reward. Rate of performance thus plays no part in determining success. This technique has revealed marked differences in behaviour between normal rats and those thyroidectomized early in life.

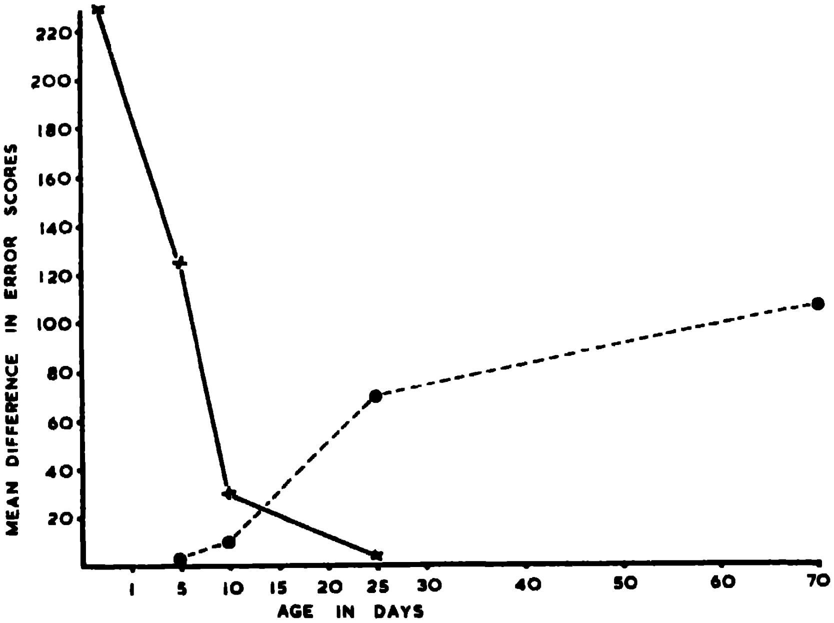


Fig. 3. Effect of thyroidectomy and of replacement therapy on adaptive behaviour as measured by the Hebb-Williams test. The ordinate records the mean increase in errors scores due to the experimental treatment: the abscissa the age at which thyroidectomy carried out or medication started. *Continuous line* thyroidectomized/unmedicated; *broken line*: thyroidectomized at birth/medicated. (Data of Eayrs, 1961.)

In the first place the cretinoid rat generally proves to be more placid and easier to handle, habituating more rapidly to the novel testing situation. Subsequently, however, not only are the error scores of rats thyroidectomized in infancy grossly inferior to those of normal littermates but the degree of impairment is correlated both with the age at which thyroidectomy is carried out and with the age at which medication is started (Fig. 3). Thus performance in the rat thyroidectomized at 24 days of age, or thyroidectomized at birth and medicated from the age of 10 days, is virtually indistinguishable from that of the normal whereas thyroidectomy earlier in life or delayed medication have a deleterious effect. These changes are not, however, confined to behaviour measured by gross error scores alone for, while the rank order of difficulty presented by the several problems has proved the same for two independent groups of normal and unmedicated thyroidectomized rats respectively it differs significantly between normal animals and littermates thyroidectomized at birth (Table 2). In this connexion it is noteworthy that this lack of correlation is directly related to the severity of behavioural impairment imposed by the several experimental treatments and may possibly be attributed to

TABLE 2. Closed-field test—effect of thyroidectomy and replacement therapy on order of difficulty presented by different problems (Eayrs, 1961)

Experimental treatment compared	Spearman Rank-order correlation coefficient	Probability that correlation absent
Two normal groups	0·893	<0·001
Thyroidectomized at birth:		
Radio vs goitrogen methods	0·818	<0·001
Normal groups vs		
Thyroidectomized at birth:	0·280	>0·10
,, 5 days	0·210	>0·10
,, 10 days	0·867	<0·001
,, 24 days	0·902	<0·001
Thyroidectomized at birth and medicated from		
5 days	0·937	<0·001
10 days	0·958	<0·001
24 days	0·601	0·05–0·02
70 days	0·346	>0·10

differences in the opportunities presented by the several problems for perseverative behaviour (see Fig. 4) or to the available sensory cues being differently used by the cretin. These factors are in need of further investigation.

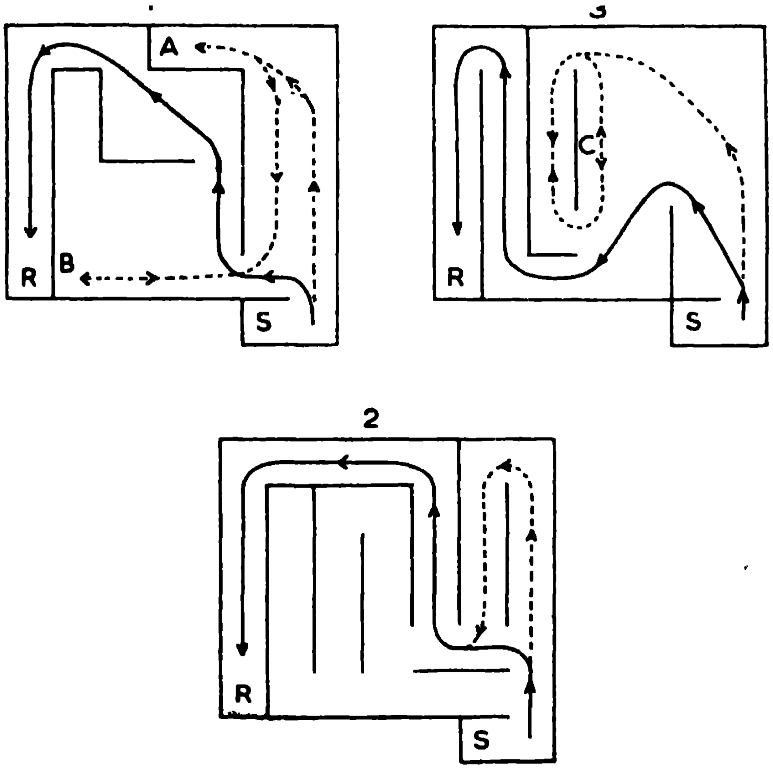


Fig. 4. Typical layouts of closed field problems affording opportunities for perseveration. Solid, direct and error-free route from start (S) to reward (R); *Broken line*: routes frequently adopted by all rats but quickly abandoned by the normal animal. Repetitive oscillation between major choice points (A-B-Set 1) or barrier circling (C-Set 3) tend to persist in cretin.

The observations of all three experiments, based as they are on littermate controls, clearly establish from the behavioural point of view that cerebral maturation is severely impaired as a result of thyroid deficiency uncomplicated by congenital factors but we are left to seek reasons for the apparently divergent effects of medication obtained in different testing situations. Evidence from a number of sources suggests that, in addition to the increased reaction time associated with hypothyroidism, the cretinoid animal is less sensitive to minor features in its environment and tends to perseverate in the performance of familiar responses. In such circumstances experimental situations where environmental cues are abundant and unchanging (such as obtained in the T-maze) or where rate of performance plays a prominent part in determining success (as in the conditioned avoidance test) may perhaps be regarded as insufficiently critical for measuring the degree of improvement resulting from medication. The closed-field test, on the other hand, where the time factor plays no part in determining success, where major cues to the solution of the problems are minimized and where opportunities for perseveration and 'carry-over' from problem to problem are present, may perhaps be regarded as providing a situation better suited to measuring the peculiar incapacities conferred by early hypothyroidism. Using this test, comparison of the success curves relating to the ages at which thyroidectomy and medication become effective suggests that the first 14 postnatal days are critical for the influence of thyroid hormone upon neural maturation—a period of development which would seem to correspond with the 4th or 5th months of gestation in man.

Electrical Activity of the Brain

The maturational influence of thyroid hormone can also be demonstrated in terms of the changes which occur in electrocortical activity (Bradley, Eayrs and Schmalbach, 1960). Records taken simultaneously through electrodes attached to the skull of normal and thyroidectomized littermates of different ages (Fig. 5) show that amplitude is diminished and the capacity to respond to environmental change either by the blocking of resting rhythms by auditory arousal or by following fixed frequencies of photic stimulation, is reduced or absent. Since these phenomena are not observed in animals in which thyroidectomy has been deferred until the 24th postnatal day it must be concluded that they reflect some change in development rather than a reduced metabolism. This conclusion is reinforced by examining the changes in the non-specific potentials evoked at the cortical surface by electrical stimulation of the medial thalamus (Bradley, Eayrs, Glass and Heath, 1961). Fig. 6 gives composite representations of these tri-phasic wave forms based upon measurements of the latency between the stimulus and beginning of the negative wave, the duration of the negative wave and its amplitude. These show in the first place that, whether thyroidectomy be carried out in infant or adult, the latency and duration of the wave are increased but can be restored to normal by giving thyroid hormone; and second that the amplitude of the wave is directly related to age, is reduced by thyroidectomy at birth but not at maturity and

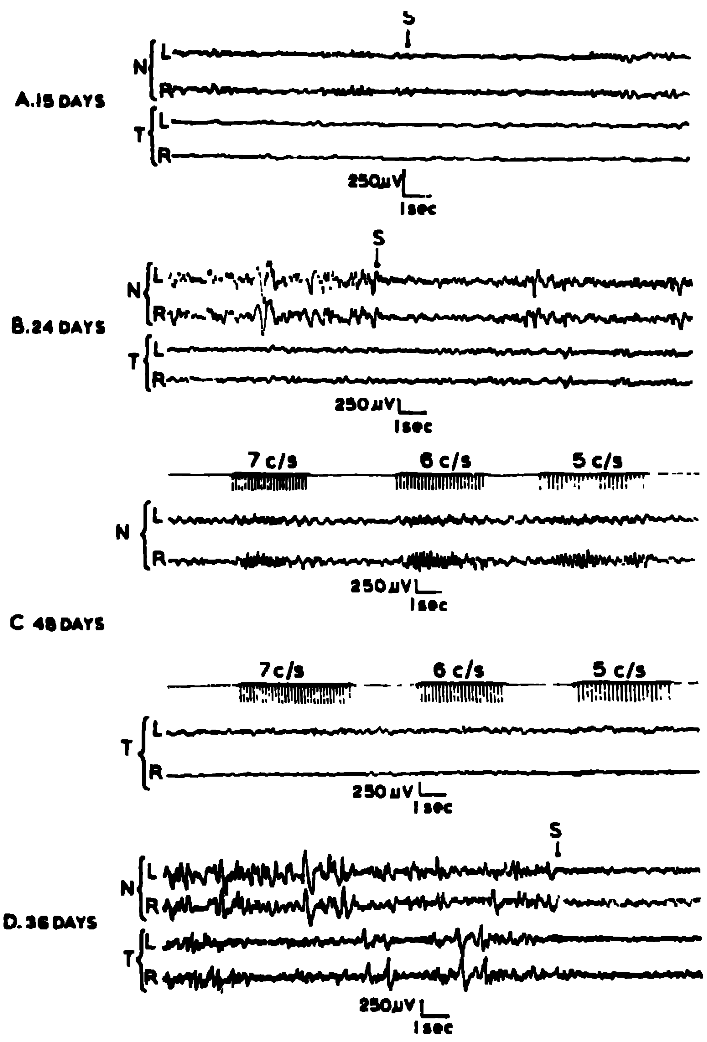


Fig. 5. Electroencephalograms taken from resting rats at the ages indicated. N: normal animal; T: littermate thyroidectomized at birth. Auditory stimulation (jangling of bunch of keys) was applied at signal (S) in records A, B and D; stroboscopic photic stimulation administered at frequencies shown in record C.

is not restored by subsequent medication. It may thus be inferred that metabolic factors do contribute to some of the properties of electrocortical activity but not to that of amplitude which, in the absence of evidence to the contrary, would seem better related to the factor of growth. One possible explanation for these findings is that the parameters of stimulation appropriate to secure an optimum response in the normal rat differ from those in the cretin. This has been investigated by varying the voltage of stimulation at a fixed frequency and by varying the frequency at a fixed voltage but the findings have afforded little reason to believe that such can be the case (Bradley, Eayrs and Richards, 1964).

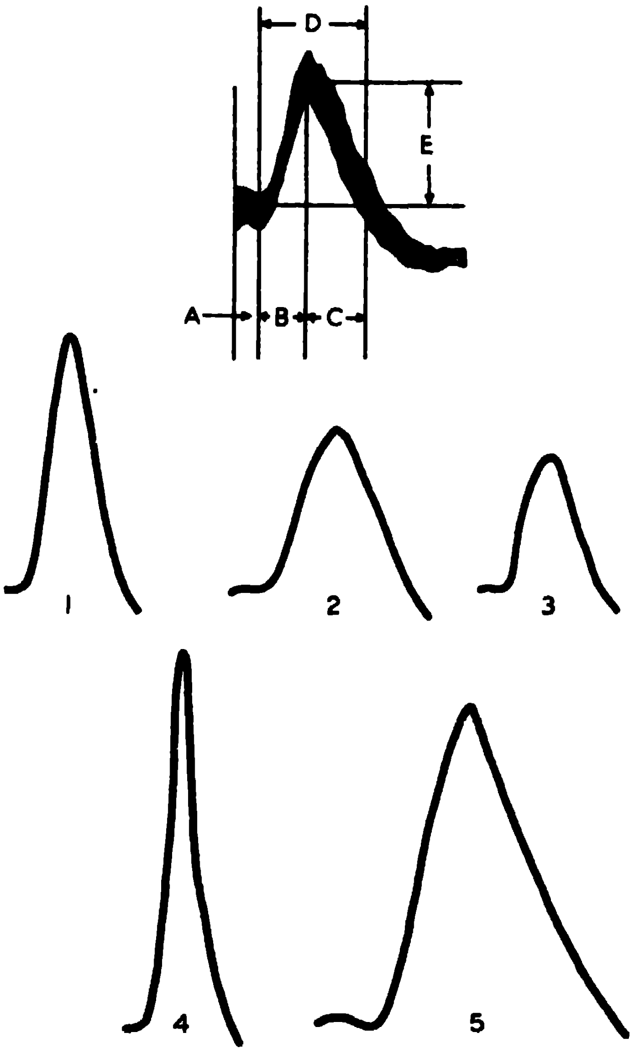


FIG. 6. Evoked potentials recorded from cortex of rat. *Top*: Characteristic record showing measurements taken. *Below*: Reconstructed potential wave-forms based on mean measurements taken from different classes of rat: 1. 25-day old normal; 2. littermate of (1) thyroidectomized at birth; 3. littermate of (2) medicated for 3 days before recording; 4. normal adult; 5. thyroidectomized adult.

Cortical Histogenesis

The effects of thyroidectomy on both behaviour and electrocortical activity in the rat have thus several features in common. In both the dichotomy between the metabolic and maturational influence of the thyroid can be detected; in both the factor of age is related to the severity of the disturbance due to thyroidectomy; and in both the administration of thyroid hormone has, in appropriate circumstances, an ameliorating influence. It is therefore of particular interest to see how far these phenomena can be explained in terms of deviation from the normal course of cerebral histogenesis. Of the many features which present themselves for examination in this respect the most significant would seem to be those which are directly concerned with the functional inter-relationship between neurones and in particular the growth of the interlacing meshwork of cell processes commonly termed neuropil.

Cortical histogenesis is characterized, in its early stages, by a stereotyped pattern of migration of neuroblasts from a mitotic zone to occupy their definitive positions in the cortical laminae in which they later differentiate to assume their adult configuration (Berry and Eayrs, 1963; Berry, Rogers and Eayrs, 1964). In the rat these events are just reaching their conclusion at the time of birth (Tilney, 1933) and any influence of postnatal thyroidectomy must thus largely be confined to the processes underlying the maturation of neurones *in situ*. Measured in that part of the cortex concerned in the mediation of the placing reflex, referred to earlier, the events which characterize this phase of development in the normal individual are an enlargement of the perikarya and a reduction in packing density which is not so much due to their increased size as to their separation by the growth of axonal and dendritic processes. These appear in part to replace material, probably glial, occupying the inter-perikaryal space and in part to take up space of their own. The growth of the axons, many of which grow in from extra-cortical origins, precedes that of the dendrites to attain a range of interlaminar densities directly related to depth beneath the pial surface of the cortex, though whether this difference in rate of growth is absolute or relative is difficult to determine in so complex a system (Eayrs and Goodhead, 1959). Comparable findings have been reported in other species, notably the guinea-pig (Peters and Flexner, 1950), mouse (Haddara, 1956) and man (Schadé, Meeter and Groeningen, 1963).

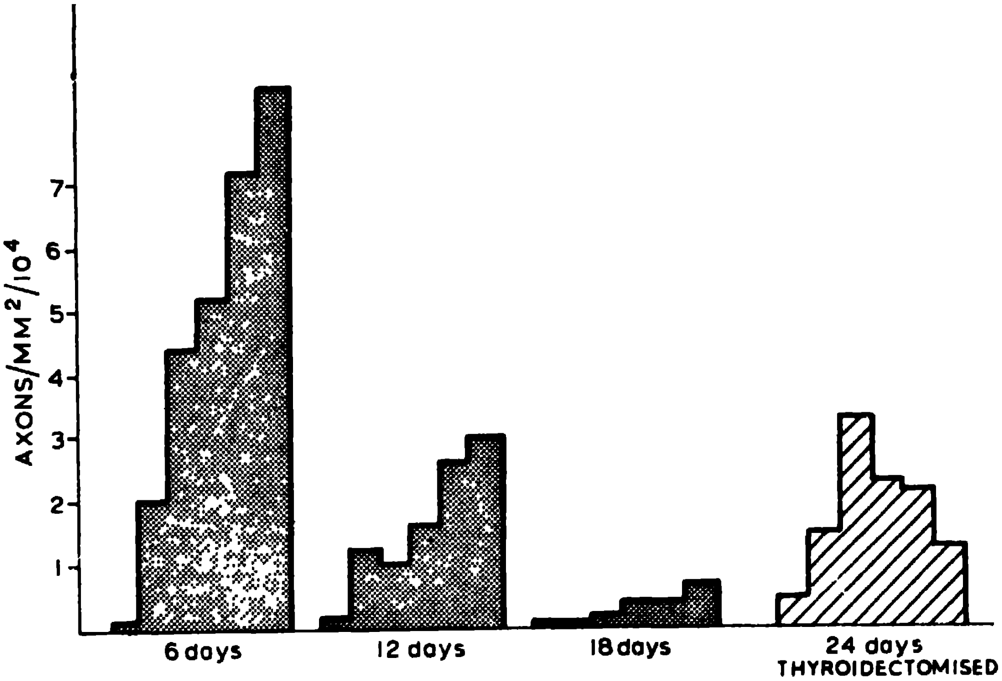


FIG. 7. Growth of axon network by laminae in cerebral cortex of rat. Dotted histograms record growth in axonal density between age shown and 24 days in cortical laminae 1, 3, 4, 5a, 5b and 6 reading from left. Hatched histogram records mean difference between normal and thyroidectomized littermates aged 24 days. Note selective effect in lamina 4.

Against this background the effects of thyroidectomy at birth are seen as a reduced rate of growth both of the perikarya and of the intervening neuropil resulting in a cortex in which the cells appear more closely aggregated (Eayrs and Taylor, 1951). The density of axons is reduced not, as might be expected were a retardation of growth alone responsible, in proportion to the densities found in the respective cortical laminae but apparently selectively in layer 4 which receives the specific afferent plexuses of axons originating in the thalamus (Fig. 7). Likewise the basal dendrites associated with each neurone are shorter and branch less with the result that the exponential mode of decay in dendritic density with distance from the centre of the perikaryon, shown to be characteristic of the normal neurone

(Sholl, 1953), is not maintained in the cretin and is in any case more rapid than in the normal (Eayrs, 1955). Combination of the data for axons and dendrites in the manner described by Uttley (1955) to provide an index of the likelihood of axodendritic interaction or connectivity shows (Fig. 8) that this factor is catastrophically reduced as early as 15 days of age and that the gap continues to widen at least up to the 36th day (Eayrs, 1960).

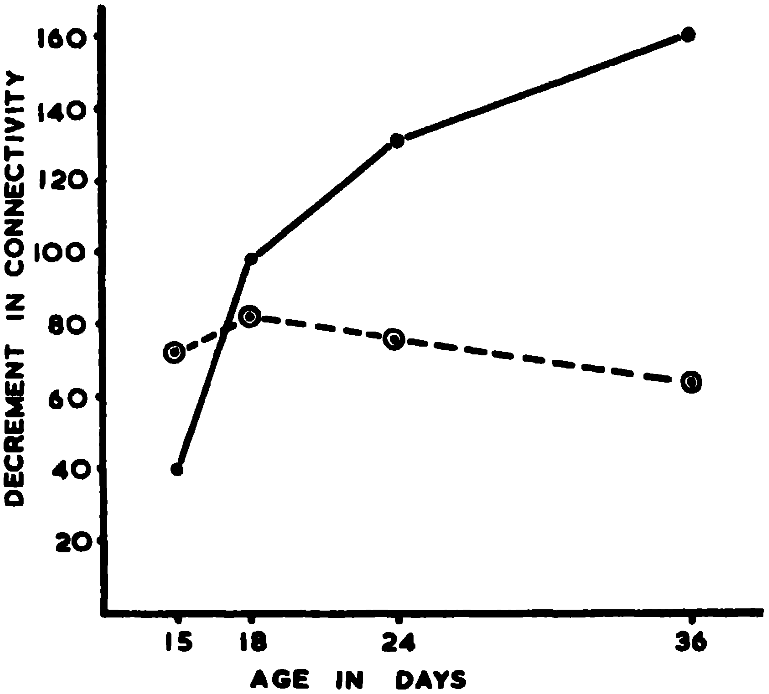


Fig. 8. Progressive decrement in axo-dendritic connectivity resulting from thyroidectomy at birth shown as absolute reduction in arbitary units (*continuous line*) and relative reduction as percentage (*broken line*).

Structural and Functional Inter-Relationship

Any attempt to inter-relate these findings, drawn as they are from three different fields of study, must in the present state of knowledge remain speculative though suggestions concerning the interplay between anatomical and physiological factors may reasonably be advanced. From the point of view of the electrical activity of the brain of the cretinoid individual the dominant changes, independent of metabolic factors, are a reduction in amplitude and an impaired response to sensory stimulation.

Both may well be related to the depletion of neuropil; for if, as has been suggested, the slow rhythmic activity of the cortex is attributable to the summation of post-synaptic dendritic potentials (Clare and Bishop, 1955; Purpura, 1959) then the changes in amplitude seen both during development and as a result of early thyroid deficiency may be explained in terms of the corresponding changes in the opportunity for axo-dendritic interaction. Likewise the failure of the electroencephalogram of the cretinoid rat to respond to sensory stimulation and the apparent insensitivity of such animals to minor environmental change may both be equated with the selective impairment to the growth of the specific afferent plexuses of the cortex. Such insensitivity may well contribute both to the poorer learning ability of the cretin and to departures from the normal in the mode of solution of problems. But a further possibility is suggested by the concept that cortical integration may be mediated as much by the modulation of the probability of all-or-none discharge of the neurone by activity within its surrounding dendritic field as by the interchange of information through axonal circuitry. The factors thought to determine this modulation are the number, arrangement and source of origin of axo-dendritic synapses (Bullock, 1959) and the density and pattern of dendritic branching (Lorente de Nó, 1960) both of which undergo developmental impairment as a result of thyroid deficiency. Removal of a considerable proportion of this modulating influence might be expected to reduce the variety of response obtainable from aggregates of neurones and thus give rise to the perseverative tendencies seen not only in the cretinoid rat but in other instances where interaction between neurones might be expected to be reduced such as, for example, in cases of asphyxia neonatorum (Becker and O'Donnell, 1952), subtotal anaesthesia (Steinberg, 1956) and cretinism in man (Money, 1956). It must be emphasized, however, that these considerations, based as they are on correlation rather than direct causal relationship, do no more than provide a working hypothesis for future study and current attempts to relate the behavioural performance of rats thyroidectomized or medicated at different ages with a measure of axo-dendritic interaction have given preliminary results which suggest the superimposed influence of factors not measured by the neuro-histological procedures adopted, such as the chemical, enzymological and ultrastructural configuration of the neurone and synapse.

DIRECT OR INDIRECT EFFECT

We may now properly inquire whether the effects we have observed are mediated directly and specifically or are the result of influences affecting the growth of the body as a whole. No satisfactory answer can yet be provided to this problem and the best that can be done is to weigh the available evidence. It has been proposed by Tusques (1955) that the effects of early hypothyroidism are secondary to malnutrition arising as a result of metabolic involvement but there is reason to doubt whether this is a sufficient explanation. In the first place, although the maturation of innate behaviour is retarded and capacity for learning later in life impaired by severe inanition during infancy, these effects are considerably less than those of thyroidectomy and to a significant extent reversible (Eayrs and Hodson-Walker—unpublished); and second, whereas a depletion of the axon network is restored by refeeding the rat starved during infancy it does not so respond to medication in the cretin (Horn, 1955). There are, however, other grounds for postulating an indirect effect. The changes, referred to earlier, in the growth of the head of the cretinoid rat are reflected in that of the brain which, by contrast with the condition in the normal animal, grows in width and height rather than in length (Eayrs and Taylor, 1951). Such changes, which may be attributable to the differential effect of thyroid hormone on endochrondral ossification at the base of the skull as opposed to that on the membranous ossification of the calvarium (Dye and Maugham, 1929), are reminiscent of those reported to result from hypophysectomy in the infant rat, a condition which is associated with a disparity in the rates of growth of brain and endocranium so extreme that the cerebellum becomes herniated through the foramen magnum (Asling, Walker, Simpson, Li and Evans, 1952). This suggests that an abnormal increase in intra-cranial pressure may occur during the course of development and, although attempts to demonstrate this experimentally have so far failed (probably by reason of unsatisfactory techniques) such a condition would be consistent with concomitant changes in the pattern of cortical vascularity (Eayrs, 1954) in which a relatively large number of small blood vessels is replaced by a smaller number of larger vessels (Fig. 9). The implication of these findings is that, while the relative blood volume of the cerebral cortex is little changed, the redistribution of the blood supply might be expected, potentially at least, to exert an influence on nervous development. Thus the overall reduction in the number of capillaries and their consequent wider spacing must alter nutritional gradients in a manner disadvantageous to the growth and maturation of neurones. Whether the impaired development of the cortical neuropil can be attributed to this factor is a moot point for evidence is lacking to show the extent to which capillarity is determined by the needs of the tissues or the capacity of tissue to develop is regulated by the available blood supply. In the former case the reduction in number of blood vessels might be expected to reflect a reduced metabolism rather than to be responsible for changes in cortical morphology and in this connexion it may be significant that the *in vitro* oxygen consumption of the cerebral cortex of the developing hypothyroid individual does not differ from normal

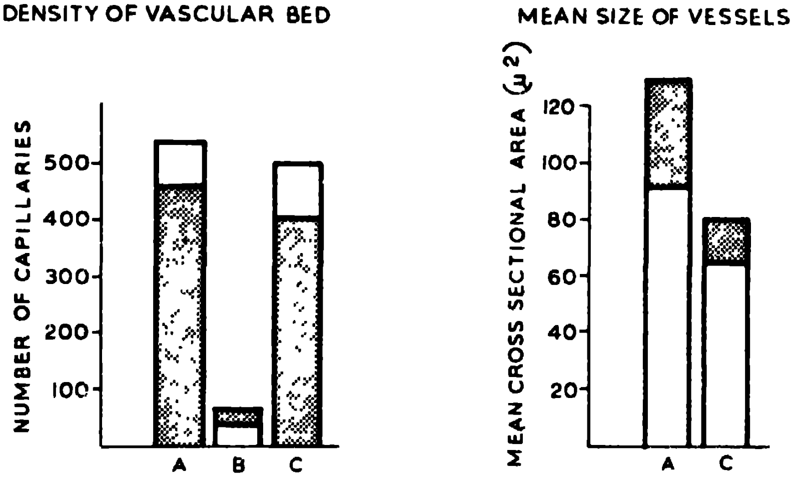


Fig. 9. Changes in cortical vascularity resulting from thyroidectomy at birth. Data for normal animal shown clear; those for the cretinous animal shown dotted. Direction of differences indicated by position of solid portion of histogram. A: all vessels; B: vessels exceeding 5^{μ} in diameter; C: vessels less than 5^{μ} in diameter.

(Fazekas, Graves and Alman, 1951) whereas an increase in the rate of blood flow in cretins medicated for a short period of time is associated with a fall rather than a rise in arterio-venous tensions (Himwich, Daly, Fazekas and Herrlich, 1942). This implies that the cerebral tissues may not take up oxygen to their capacity and a possible explanation of this phenomenon is that the capillary bed is inadequate to transport an increased oxygen supply to the tissues. The shift in the size of blood vessels, however, has an added significance in so far as it may arise from the restriction in the venous return attributable to an increased intra-cranial pressure thus giving rise to a condition of stagnant anoxia.

While it is probable that such indirect factors play a significant part in determining the effects of thyroid deficiency during development there is good reason to believe that thyroid hormone exerts a more direct influence upon central nervous maturation. In the amphibian nervous system, for instance, there can be no doubt that the hormone is an important determinant of the reorganization needed to equip the metamorphosed individual for a new mode of life (Weiss and Rossetti, 1951). In the mammal, if the rat be regarded as a typical example, the changes in central nervous morphology described earlier can, to a large extent, be attributed to an impairment in the mechanism of the synthesis of protein and it is therefore of considerable interest to know to what extent this depends on the activity of the thyroid. Hamburgh and Flexner (1957) demonstrated that the development of at least one enzyme system, succinic dehydrogenase, was irreversibly impaired by neonatal thyroidectomy if therapy were deferred until the 15th day of age. More recent work has thrown new light on this problem. This has shown that the incorporation of amino acid into protein, which proceeds considerably faster in immature than in adult brain, is accelerated by the addition of thyroxine to preparations of the former but not of the latter tissue (Gelber, Campbell, Diebler and Sokoloff, 1964). This difference of effect has been shown to reside in dissimilarities in the properties of the mitochondria (Klee, Cason and Sokoloff, 1963) and it is thus possible that maturational changes in these structures may account for the failure of thyroid therapy after a critical age.

Results, recently published by Kidson and Kirby (1964), have also specifically implicated the action of thyroid hormone in the synthesis of messenger RNA and these combined findings could ultimately prove of the greatest importance in supplementing an understanding of the causes of the poor intellectual abilities of the cretin. An impaired capacity to synthesize protein could clearly be responsible in the first place for the reduced opportunity for interaction between neurones and for modulation of axonal discharge to which reference has already been made. There has of late, however, been a tendency to pay greater attention to so-called molecular theories of learning and memory which regard the memory store as inherent in the coding of the RNA molecule. Detailed discussion of these theories (see Dingman and Sporn, 1964; Rose, 1964) which, in any case, are but vaguely formulated at the present time falls outside the scope of this communication. They have been mentioned largely to draw attention to the possible linkage between RNA metabolism, thyroid hormone and memory processes which may play an important part in future thinking on the subject.

CONCLUSION

It may be concluded from this survey that thyroid hormone exerts an important influence on the maturation of the mammalian brain. There is so far no evidence to suggest that it plays any part in organizing the centres mediating innately organized behaviour but its presence would appear to be essential for a proper development of cortical structure and function. Possibly operating directly through its influence on the synthesis of protein or indirectly through an effect on the cerebral environment, uncomplicated thyroid deficiency arising before a critical stage of development gives rise to irreversible changes in cerebral maturation. It seems likely that this critical period which, in the rat, occurs between the 10th and 15th day post-partum occurs during the foetal life of man and that inconsistencies in the response of the human cretin to post-natal medication may reflect the availability of thyroid hormone to the developing individual during this period.

REFERENCES

Asling, C. W., Walker, D. G., Simpson, M. E., Li, C. H. and Evans, H. M. (1952) *Anat. Rec.*, **114**, 49.

Becker, R. F. and O'Donnell, W. (1952) *J. comp. physiol. Psychol.*, **45**, 153.

Berry, M. and Eayrs, J. T. (1963) *Nature Lond.*, **197**, 984.

Berry, M., Rogers, A. W. and Eayrs, J. T. (1964) *Nature Lond.*, **203**, 591.

Bradley, P. B., Eayrs, J. T. and Schmalbach, K. (1960) *Electroenceph. clin. Neurophysiol.*, **12**, 467.

Bradley, P. B., Eayrs, J. T. and Richards, N. M. (1964) *Electroenceph. clin. Neurophysiol.*, **16**, 622.

Bradley, P. B. Eayrs, J. T., Glass, A. and Heath, R. W. (1961) *Electroenceph. clin. Neurophysiol.*, **13**, 577.

Brown, A. W., Bronstein, I. P. and Kraines, R. (1939) *Amer. J. Dis. Child.*, **57**, 517.

Bruch, H. and McCune, D. J. (1944) *Amer. J. Dis. Child.*, **67**, 205.

Bullock, T. H. (1959) *Science*, **129**, 997.

Clare, M. C. and Bishop, G. H. (1955) *Amer. J. Psychiat.*, **111**, 818.

Dingman, W. and Sporn, M. B. (1964) *Science*, **144**, 26.

Dye, J. A. and Maugham, G. H. (1929) *Amer. J. Anat.*, **44**, 331.

Eayrs, J. T. (1954) *J. Anat. Lond.*, **88**, 164.

Eayrs, J. T. (1955) *Acta anat.*, **25**, 160.

Eayrs, J. T. (1960) In *Structure and function of the cerebral cortex*, ed. Tower, D. B. and Schadé, J. P. Elsevier: Amsterdam.

Eayrs, J. T. (1961) *J. Endocrin.*, **22**, 409.

Eayrs, J. T. (1964) *Anim. Behav.*, **12**, 195.

Eayrs, J. T. and Goodhead, B. (1959) *J. Anat. Lond.*, **93**, 38.

Eayrs, J. T. and Lishman, W. A. (1955) *Brit. J. Anim. Behav.*, **3**, 17.

Eayrs, J. T. and Holmes, R. L. (1964) *J. Endocrin.*, **29**, 71.

Eayrs, J. T. and Levine, S. (1963) *J. Endocrin.*, **25**, 505.

Eayrs, J. T. and Taylor, S. H. (1951) *J. Anat. Lond.*, **85**, 350.

Etkin, W. (1955) 'Metamorphosis'. In *Analysis of Development*, ed. Willier, B. H., Weiss, P. A. and Hamburger, V. Saunders: Philadelphia.

Evans, H. W., Simpson, M. E. and Pencharz, R. I. (1939) *Endocrinology*, **25**, 175.

Faxen, N. (1935) *Acta paediat.*, **17**, 565.

Fazekas, J. F., Graves, L. B. and Alman, R. W. (1951) *Endocrinology*, **48**, 169.

Gelber, S., Campbell, P. L., Diebler, G. E. and Sokoloff, L. (1964) *J. Neurochem.*, **11**, 221.

Goldberg, R. C. and Chaikoff, I. L. (1949) *Endocrinology*, **45**, 64.

Gudernatsch, J. F. (1912) *Arch. Enter. Mech. Org.*, **35**, 457.

Haddara, M. A. H. (1956) *J. Anat. Lond.*, **90**, 494.

Hamburgh, M. and Flexner, L. B. (1957) *J. Neurochem.*, **1**, 279.

Hamburgh, M. and Vicari, E. (1957) *Anat. Rec.*, **127**, 302.

Himwich, H. E., Daly, D.. Fazekas, J. F. and Herrlich, H. C. (1942) *Amer. J. Psychiat.*, **98**, 489.

Horn, G. (1955) *Anat. Rec.*, **121**, 63.

Hughes, A. M. (1944) *Endocrinology*, **34**, 69.

Jackson, A. S. (1934) *West. J. Surg.*, **42**, 498.

Kidson, C. and Kirby, K. S. (1964) *Nature, Lond.*, **203**, 599.

Klee, C. B., Cason, J. and Sokoloff, L. (1963) *Fed. Proc.*, **22**, 580.

Kollross, J. J. (1943) *Physiol. Zool.*, **16**, 269.

Lewis, A. (1937*a*) *Lancet*, **i**, 1505.

Lewis, A. (1937*b*) *Lancet*, **ii**, 5.

Lorente de Nó, R. (1960) In *Structure and function of the cerebral Cortex*, ed. Tower, D. B. and Schadé, J. P. Elsevier: Amsterdam.

McGirr, E. M. and Hutchison, J. H. (1955) *J. clin. Endocrin.*, **15**, 668.

Means, J. H. (1948) *The Thyroid and its Diseases*. Lippincott: Philadelphia.

Money, J. (1956) *Arch. Neurol. Psychiat. Chic.*, **76**, 296.

Osorio, C. and Myant, N. B. (1960) *Brit. med. Bull.*, **16**, 159.

Peters, V. B. and Flexner, L. B. (1950) *Amer. J. Anat.*, **86**, 133.

Pickering, D. E. and Fisher, D. A. (1958) *J. chron. Dis.*, **7**, 242.

Pickering, D. E. and Koulischer, N. (1956) *Amer. J. Dis. Child.*, **92**, 63.

Purpura, D. (1959) In *International Reviews of Neurobiology*, ed. Pfeiffer C. C. and Smithies, J. R. Vol. 1. Academic Press: New York.

Rabinovitch, M. S. and Rosvold, H. E. (1951) *Canad. J. Psychol.*, **5**, 122.

Radwin, L. S., Michelson, J. P., Kramer, B. and Berman, A. B. (1949) *Amer. J. Dis. Child.*, **78**, 821.

Ray, R. D., Simpson, M. E., Li, C. H., Asling, C. W. and Evans, H. W. (1950) *Amer. J. Anat.*, **86**, 479.

Rose, W. (1964) *Discovery*, **25**, 38.

Schadé, J. P., Meeter, K. and Van Groeningen, W. B. (1963) *Acta morph. neerland-scand.*, **5**, 37.

Scow, R. O., Simpson, M. E., Asling, C. W., Li, C. H. and Evans, H. M. (1949) *Anat. Rec.*, **104**, 445.

Sholl, D. A. (1953) *J. Anat. Lond.*, **87**, 387.

Smith, D. W., Blizzard, R. M. and Wilkins, L. (1957) *Pediatrics*, **19**, 1011.

Steinberg, H. (1956) *Brit. J. Psychol.*, **47**, 183.

Tilney, F. (1933) *Bull. neurol. Inst. N.Y.*, **3**, 252.

Tredgold, A. F. (1929) *Mental Deficiency*, 5th edn. Ballière, Tindall and Cox: London.

Tusques, J. (1955) *Biologie (Paris)*, **45**, 395.

Uttley, A. M. (1955) *Proc. roy. Soc. B.*, **144**, 229.

Walker, D. G., Asling, C. W., Simpson, M. E., Li, C. H. and Evans, H. W. (1952) *Anat. Rec.*, **114**, 19.

Weiss, P. and Rossetti, F. (1951) *Proc. nat. Acad. Sci. Wash.*, **37**, 540.

Werner, S. C. (1962) *The Thyroid: a fundamental and clinical text*. 2nd edn. Hoeber: New York.

INDEXES

'*The Scientific Basis of Medicine*'

INDEX OF AUTHORS

Note. Roman numerals refer to the eight volumes published under the title *Lectures on the Scientific Basis of Medicine*. Subsequent volumes published as *The Scientific Basis of Medicine Annual Reviews* are referred to by the distinguishing year date of their title.

Abercrombie M., VIII, 19
Adrian, E. D., I, 1
Albert, Adrien, II, 14
Andrewes, C. H., IV, 18; *1962*, 20
Armstrong, C. N., VIII, 48
Astbury, W. T., VIII, 429

Barber, Mary, VI, 333; *1964*, 169
Barnes, J. M., V, 244
Bates, J. A. V., V, 333
Bauer, D. J., *1966*, 174
Bedson, S. P., I, 214
Belcher, E. H., *1964*, 250
Bergel, F., IV, 54
Biggs, Rosemary, VII, 80
Billing, Barbara H., *1963*, 197
Bishop, P. M. F., III, 169
Black, D. A. K., *1964*, 291
Booth, C. C., *1963*, 171
Born, G. V. R., *1963*, 249
Boyland, E., III, 327
Brown, J. McClure, V, 90
Brown, Sir Lindor, *1961*, 1
Bülbring, Edith, VII, 374
Bull, G. M., III, 219
Bull, J. P., IV, 167
Buller, A. J., *1965*, 186
Burn, J. H., III, 127
Byrom, F. B., VIII, 256

Cameron, Sir Roy, I, 130; III, 52; *1961*, 324
Cappell, D. F., VIII, 329
Carmichael, E. A., III, 93
Case, R. A. M., *1966*, 71
Catcheside, D. G., *1962*, 140
Causey, G., *1962*, 294
Chesterman, C. C., III, 314
Clark, Sir Wilfrid le Gros, V, 1
Cockcroft, Sir John, VI, 1
Cohen of Birkenhead, Lord, *1963*, 1
Collier, L. H., *1962*, 345
Cooper, K. E., *1965*, 239
Cope, C. L., VIII, 306
Court-Brown, W. M., *1962*, 109
Cross, K. W., VII, 348
Cumings, J. N., *1965*, 37
Cuthbertson, W. F. J., V, 285

Dacie, J. V., VII, 59
Dalgliesh, C. E., *1961*, 129
Daniel, P. M., *1963*, 83
Davidson, J. N., IV, 34; *1965*, 17
Davies, H. G., V, 105
Davson, H., *1966*, 238
Dawes, G. S., V, 53
Dempster, W. J., VI, 13
Dent, C. E., II, 213
Dick, G. W. A., VII, 426
Dodds, Sir Charles, Bt., *1965*, 1
Dodds, E. C., I, 285
Doll, W. R. S., *1961*, 74
Doniach, D., *1965*, 110
Donovan, B. T., *1963*, 53
Dornhorst, A. C., V, 398
Douglas, J. W. B., *1966*, 91
Downie, A. W., I, 230

Eayrs, J. T., *1966*, 317
Edholm, O. G., I, 79; V, 36
Evans, D. G., VIII, 132

Feldberg, W. S., II, 343; *1963*, 137
Fell, Honor B., VI, 28
Fleming, Sir Alexander, II, 29
Fourman, P., VII, 451; *1961*, 270
Frazer, A. C., IV, 311
Fridemann, I., *1963*, 302

Gaddum, J. H., IV, 238
Galton, D. A. G., *1962*, 152
Garnham, P. C. C., II, 323
Garrod, L. P., III, 302
Gell, P. G. H., *1964*, 189
Gilson, J. C., VI, 58
Glynn, I. M., *1966*, 217
Glynn, L. E., VII, 359; *1962*, 94
Goffe, Alan, *1964*, 151
Goldblatt, M. W., V, 262
Gordon, G., VII, 398
Gray, C. H., IV, 74; *1961*, 200
Gray, J. A. B., VII, 415
Gray, L. H., VII, 314
Greenbaum, A. L., VI, 277
Greenfield, A. D. M., VIII, 181

Hadfield, G., VI, 297
Haldane, J. B. S., III, 41
Hallpike, C. S., VII, 1
Harkness, R. D., V, 183; *1961*, 236
Harris, G. W., III, 154
Harrison, A. R., *1963*, 266
Harrison, C. V., VIII, 226
Harrison, R. G., *1963*, 76
Hayes, W., VIII, 107
Hewitt, H. B., *1962*, 305
Hill, A. V., IV, 1
Hilton, S. M., *1965*, 217
Hobbs, J. R., *1966*, 106
Horning, E. S., V, 421
Huggett, A. St. G., II, 235
Humphrey, J. H., VII, 239; VIII, 87
Hunt, J. N., V, 348

Isaacs, A., *1961*, 21

James, A. T., *1961*, 262
Jefferson, Sir Geoffrey, I, 61
Jenkins, G. N., VIII, 442
Johnson, F. R., *1964*, 276
Joseph, J., *1963*, 158
Judah, J. D., IV, 224

Kay, H. D., I, 39
Kay, H. E. M., *1962*, 327
Keele, C. A., VI, 143
Kekwick, R. A., IV, 112
Kellgren, J. H., III, 78
King, E. J., II, 108; III, 187
Klyne, W., *1964*, 317
Knox, R., *1962*, 77
Krebs, H. A., V, 126
Krohn, P. L., VII, 285

Lajtha, L. G., *1961*, 63
Lane-Petter, W., *1966*, 54
Lennox, B., *1962*, 124
Lewis, A. J., VI, 116
Lewis, G. P., *1962*, 242
Liddell, E. G. T., VI, 100
Linden, R. J., *1965*, 164
Lindop, Patricia J., *1965*, 91
Long, C., VI, 217
Longmuir, I. S., *1966*, 300
Loutit, J. F., I, 379; V, 439
Lowbury, E. J. L., *1962*, 57

McClean, D., IV, 266
Macfarlane, R. G., I, 244
McIlwain, H., VI, 168
Macintyre, I., *1963*, 216
MacKenna, R. M. B., III, 347
Mackintosh, J. M., I, 267
Magee, P. N., *1962*, 172
Maitland, H. B., III, 259
Maizels, M., VIII, 287
Marrian, G. F., II, 182
Marshall, John, *1964*, 67
Martin, N. H., V, 165
Mason, A. Stuart, *1965*, 259
Mayneord, W. V., *1966*, 16
Melrose, D. G., IV, 252; VI, 85
Mercer, E. H., VIII, 32
Michael, R. P., *1965*, 316
Miles, A. A., I, 192; III, 235; VIII, 198
Millen, J. W., *1963*, 125
Miller, J. F. A. P., *1964*, 218
Mills, J. N., *1966*, 192
Milne, M. D., V, 404; *1961*, 149
Mitchison, D. A., *1963*, 319
Mole, R. H., VIII, 65

Mollin, D. L., vii, 94
Mollison, P. L., ii, 269
Montgomery, G. L., vii, 162
Morgan, W. T. J., iv, 92
Morton, R. A., v, 143
Moss, D. W., *1964*, 334
Myant, N. B., vi, 313

Nabarro, J. D. N., *1962*, 277
Negus, Sir Victor, iv, 332
Neuberger, A., ii, 195
Nordin, B. E. C., 308, *1964*

Oswald, Ian, *1964*, 102

Parker, M. T., *1966*, 157
Parkes, A. S., ii, 250; *1963*, 25
Paton, W. D. M., ii, 139; iii, 105
Peacock, P. R., ii, 68
Pearse, A. G. Everson, iv, 358; *1962*, 215
Peart, W. S., vii, 182
Penrose, L. S., vii, 25
Pepys, J., *1964*, 203
Perkins, F. T., *1965*, 282
Perry, S. V., v, 314
Phillips, C. G., *1964*, 81
Pickering, Sir George, *1961*, 97
Pickford, L. Mary, iv, 286
Platt, B. S., iv, 145
Platt, Sir Robert, *1966*, 1
Polani, P. E., *1965*, 141
Pollock, M. R., *1962*, 34
Pontecorvo, G., *1961*, 13
Poole, J. C. F., *1964*, 55
Popjak, G., iii, 17
Prankerd, T. A. J., viii, 269
Pritchard, J. J., *1963*, 286
Pugh, L. G. C. E., *1964*, 32

Rees, R. J. W., vii, 203
Reid, Lynne, viii, 235
Rob, C. G., vi, 46
Robb-Smith, A. H. T., ii, 77
Roberts, J. A. Fraser, vii, 42
Roddie, I. C., *1966*, 260
Roitt, I. M., *1965*, 110
Rosenheim, M. L., i, 96
Rowles, S. L., *1962*, 259
Rudall, K. M., v, 217; *1962*, 203
Russell, R. W., iii, 372
Russell-Fraser, T., *1963*, 36

Salisbury, Sir Edward, viii, 1
Saunders, L., *1965*, 270
Scowen, E. F., viii, 358
Seddon, H. J., vi, 363
Shaldon, Stanley, *1966*, 201
Shepherd, M., *1965*, 299
Sherlock, Sheila, *1961*, 216
Shillingford, J., vii, 472
Sigerist, H. E., iii, 1
Sinclair, H. M., vi, 260
Slome, D. A., v, 451
Smith, Audrey U., v, 19
Smith, C. E. Gordon, *1964*, 125
Smith, K. M., vi, 379
Smith, Wilson, ii, 46
Spence, Sir James, ii, 1
Squire, John R., iv, 123; *1966*, 128
Stamp, Lord, iii, 279
Stevenson, Alan C., *1964*, 234
Stewart, W. K., ii, 334
Stoker, M. G. P., vii, 223
Stokes, J. F., v, 380
Stoner, H. B., *1961*, 172
Strang, L. B., *1965*, 202
Stuart-Harris, C. H., viii, 148
Symington, *1964*, 15

Tanner, J. M., i, 308
Ten Cate, A. R., *1966*, 146
Thomlinson, R. H., *1965*, 74
Thompson, R. H. S., ii, 165; *1961*, 283; *1965*, 58
Todd, Lord, *1964*, 1
Tunbridge, R. E., v, 231

Van Slyke, D. D., i, 143
Vaughan, Dame Janet, iv, 196; *1961*, 47

Waddington, C. H., *1962*, 1
Walker, J., v, 67
Walls, E. W., *1963*, 113
Walshe, J. M., viii, 407
Warbrick, J. G., *1963*, 99
Wells, P. N. T., *1966*, 38
West, John, *1966*, 279
Whelan, W. J., vi, 233
White, J. C., ii, 287
White, R. G., *1961*, 31
Whittaker, V. P., vi, 186
Whitteridge, D., iv, 306
Wilkie, D. R., v, 297

Wilkinson, J. H., *1961*, 108
Williams, E. M. Vaughan, *1961*, 302
Williams, R. E. O., VIII, 165
Wilson, G. M., IV, 182
Woodruff, M. F. A., *1965*, 124
Wootton, I. D. P., *1963*, 235
Wright, G. Payling, VII, 263
Wrong, O. M., VIII, 386

Yoffey, J. M., VII, 127
Young, F. G., I, 16; III, 137
Yudkin, J., I, 364

Zaimis, Eleanor J., VI, 208

'*The Scientific Basis of Medicine*'

INDEX OF TITLES

Note. **Papers are indexed under the principal word or words of their titles. Roman numerals refer to the eight volumes published under the title *Lectures on the Scientific Basis of Medicine*. Subsequent volumes published as *The Scientific Basis of Medicine Annual Reviews* are referred to by the distinguishing year date of their title.**

Acetylcholine: The Action of Quinidine, Acetylcholine and Anaphylaxis interpreted from Simultaneous Records of Contractions and Intracellular Potentials in the Heart, *1961*, 302
— and the Maintenance of the Cardiac Rhythm, iii, 127
— in Nervous Tissue, The Metabolism of, vi, 186
Acid-base Balance, Renal Control of, v, 404
Adaptation: Physiological Adaptation in Micro-organisms, *1962*, 34
Adenoviruses and Respiratory Disease in Man, The, viii, 148
Adolescence, Growth of the Human at the Time of, i, 308
Adrenal Activity in Man, The Measurement of, viii, 306
— Cortex: The Pathological Physiology of the Human Adrenal Cortex and its Relationship to Hypercorticalism, *1964*, 15
— Hormones and ACTH, i, 16
Adrenocortical Hormones, The Metabolism of the, ii, 182; *see also* Adrenal Cortex
Ageing, Research in, i, 285; *see also* Female Reproductive Tract
Alcohol, The Effects of, iv, 238
Alimentary Tract and Body-Fluid, *1964*, 291
Altitude, High, Man at, *1964*, 32
Aminoacid Metabolism, Biochemical Aspects of Disorders of, *1961*, 129
—, Chromatography in the Study of, ii, 213
—, Disorders of, *1961*, 149
Anaemias, The Megaloblastic, vii, 94
Anaphylaxis: The Action of Quinidine, Acetylcholine and Anaphylaxis interpreted from Simultaneous Records of Contractions and Intracellular Potentials in the Heart, *1961*, 302
Anatomy, Anorectal, *1963*, 113
—: Some Aspects of the Anatomy of the Pia Mater and the Choroid plexuses, *1963*, 125
Animals: Factors in the Transmission of Virus infections from Animals to Man, *1964*, 125
—, Sophisticated Laboratory, *1966*, 54
Anorectal Anatomy, *1963*, 113
Anoxia of the New-born, vii, 348
—: Physiological Effects of Anoxia in the Foetal and Newborn Lamb, v, 53
—: Tumour Anoxia and the Response to Radiation, *1965*, 74
—: *see also* The Oxygen Environment of the Human Foetus
Antibiotic Therapy, Causes of Failure in, iii, 302
Antibiotics, Recent Progress in, ii, 29
Anticholinesterases, iii, 105
—, Cholinesterases and, ii, 165
Antimalarial Drugs, iii, 314; *see also* Malaria Parasite
Antiviral Action of Interferon, The, *1961*, 21
— Chemotherapy, *1966*, 174
— Immunity, iii, 259
Aromatic Compounds: Retention of Aromatic Compounds in Acute Renal Failure, *1963*, 235
Arterial Substitutes, vi, 46
Atomic Energy, The Biological Significance of, vi, 1
Autoimmunity and Disease, *1965*, 110
— and Diseases of the Connective Tissue, *1962*, 94
Autonomic Nervous System, The Physiology of the, ii, 343

Bacterial Enzymes: The Action of Bac-

terial Enzymes on Immunizing Antigens, iii, 279
Bacterial Invasion, Reactions to, iii, 235
Behaviour, Neurological Mechanisms and Control of Sexual, *1965*, 316
Bile Pigment Metabolism, *1963*, 197
Biochemical Abnormalities, The Genetics of Some, iii, 41
Biochemistry in Medicine, The Role of, viii, 287
Biological Synthesis, iii, 17
Biology: The Nature and Importance of Developmental Biology, *1962*, 1
Biophysics: Why Biophysics?, iv, 1
Birth, Lungs at, *1965*, 202
Blenorrhoea, Trachoma and Inclusion, *1962*, 359
Blood Coagulation in Theory and Practice, i, 244
Blood Enzymes in Diagnosis, *1961*, 108
Blood Flow, Nervous Control of Limb, *1966*, 260
—, Pulmonary, Distribution of, *1966*, 279
Blood Group Genetics, vii, 42
— Specificity in Man, The Chemical Basis of, iv, 92
Blood Platelets, Functions of the Adenine Nucleotides of, *1963*, 249
Blood Pressure, Lability of, i, 96
Blood Vessels in the Limbs, The Regulation of the, viii, 181
Blood-Bone Equilibrium, The, *1964*, 308
Body Water Control, iii, 219; *see also* Body-Fluid, The Alimentary Tract
Body-Fluid, The Alimentary Tract and, *1964*, 291
Bone: The Blood-Bone Equilibrium, *1964*, 308
—, Radiation Injury to, iv, 196
— Healing, *1963*, 286
— Marrow Studies, Autoradiography in, *1961*, 63
Botany: The Influence of Botany on Health and Disease, viii, 1
Bradykinin—Biochemistry, Pharmacology, and its Physiological Role in controlling Local Blood Flow, *1962*, 242
Brain, An Unorthodox Pharmacological Approach to the, *1963*, 137
— Trace Elements in the, *1965*, 37
Breast Cancer, The Role of the Endocrine Systemin, vi, 297
Bronchitis: Chronic Bronchitis and Hypersecretion of Mucus, iii, 235
Burns, Shock from, iv, 167

Cajal and Sherrington, vi, 100
Cancer, Chemotherapy of, iii, 327
—, Fundamental Aspects of the Radiotherapy of, *1962*, 305
Carcinogenesis, ii, 68
—, Some Anomalies in Endocrine, v, 421; *see also* Breast Cancer
Cardiovascular Responses: Hypothalamic Control of the Cardiovascular Responses in Fear and Rage, *1965*, 217
Cell, Electron Microscopy and the Living, viii, 32
— Membranes, Interfacial Films and, *1965*, 270
— Regeneration as seen in the Liver Parenchyma, vii, 263
— Surface, The Control of Growth and the, viii, 19
Cells: Preservation of Living Cells at Low Temperatures, ii, 250
—, Transplantation of Haemopoietic, *1962*, 327
Central Nervous Development, Thyroid and, *1966*, 317
Cerebral Circulation with special reference to Strokes, The, *1964*, 67
Cerebrospinal Fluid, Formation and Drainage, *1966*, 238
Chemotherapy, Antiviral, *1966*, 174
—: Contributions of Chemotherapy to the Study of Leukaemia, *1962*, 152
Cholinesterases and Anti-Cholinesterases, ii, 165; *see also* Anticholinesterases
Choroid Plexuses, Some Aspects of the Anatomy of the Pia Mater and the, *1963*, 125
Christmas Disease, Haemophilia and, vii, 80
Chromatography in the Study of Amino-Acid Metabolism, ii, 213
Chromosome Analysis and Clinical Medicine, *1962*, 109
— Anomalies, Nuclear Sexing and Human, *1962*, 124

Circulation: The Cerebral Circulation with special reference to Strokes, *1964*, 67
— The Study of the Circulation by Dye Dilution Curves, vi, 472
Clinical Science, The Methodology of, ii, 1; *see also* Medical Research, Scientific Research
Cold: The Effects of Cold on Man, v, 36; *see also* Hypothermia
Collagen, Metabolism of, v, 183
—: The Structure and Stability of Collagen in Relation to Diseases of Connective Tissue, vii, 359
Collagen-apatite Partnership, Fine Structural Studies of the, viii, 429
Common Cold, Recent Work on the, *1962*, 20
Connective Tissue, Autoimmunity and Diseases of the, *1962*, 94
—, The Functional Significance of, ii, 77
— Fibres, Observations on the Structure of, v, 231
Constitution: Human Constitution and Disease, *1963*, 1
Cooling of the Whole Organism, iv. 252; *see also* Hypothermia
Coronary Artery Disease, Problems in the Pathology of, vii, 162
Cortex: Observations on the Excitable Cortex in Man, v, 333
Cortical Mechanisms, On the Organization of, i, 61

Demography and the Cancers, *1966*, 71
Demyelinating Diseases, *1965*, 58
Dental Caries, The Biochemical Background to the Action of Fluoride in, viii, 442
Dental Tissues, Cell Biology, *1966*, 146
Dentistry, Endocrine Aspects of, *1965*, 259
Dermatology, The Scientific Approach to, iii, 347
Development: Aspects of Facial and Nasal Development, *1963*, 99
Diabetes Mellitus, Non-pancreatic aspects of, *1961*, 200
Dialysis, Long-term, as Substitute for Kidney Function, *1966*, 201
Dreaming, Physiology of Sleep Accompanying, *1964*, 102
Drug Resistance and the Response to Therapy, *1962*, 57
—: Contrasting Patterns of Drug Resistance in Tubercle Bacilli and Staphylococci, *1962*, 77

Ear, Nose and Throat, Electron Microscopy in the Pathology of, *1963*, 302
Electromyography, Posture and, *1963*, 158
Electron Microscopy in the Pathology of the Ear, Nose and Throat, *1963*. 302
Enamel, Biochemistry of, *1962*, 259
Endocrine Aspects of Dentistry, *1965*, 259
Enzyme Histochemistry and its Biochemical Alignments, *1962*, 215
Enzymes: Blood Enzymes in Diagnosis, *1961*, 108
Epidemiological Methods in Leukaemia Research, *1961*, 74
Epidemiology and Mental Disorder, *1965*, 299
— of Tubercle Bacilli, The, *1963*, 319
Epithelium: Reaction of Epithelium to Injury, *1964*, 276
Ethical Basis of Medical Science, *1966*, 1
Extracellular Fluid Volume, Sodium Excretion and the Control of, viii, 386

Facial and Nasal Development, Aspects of, *1963*, 99
Fat Metabolism, iv, 311
—, The Control of, vi, 277
Fatty Acids: Essential Fatty Acids in Nutrition and their relation to other Vitamins, vi, 260
Fear and Rage, Hypothalamic Control of the Cardiovascular Responses in, *1965*, 217
Female Reproductive Tract, Ageing Process in the, vii, 285
Fever and Pyrogens, *1961*, 97
Foetal and Newborn Lamb, Physiological Effects of Anoxia in the, v, 53
Foetus, The Oxygen Environment of the Human, v, 67; *see also* Anoxia

Gamma Globulins, Metabolism of, vii, 239

Gamma Globulins: What are Gamma Globulins?, viii, 87
Ganglionic Block, The Principles of, ii, 139
Gastric Digestive Function in Man, The Investigation of, v, 348
Genes, Functional Structure of, *1962*, 140
Genetics: Bacterial Genetics and Gene Structure, viii, 107
—: Biochemical Genetics, ii, 195
—: Biochemical Genetics and Medicine, vii, 25
—: Genetic Analysis via Somatic Cells, *1961*, 13
—: Radiation-Induced Genetic Changes, *1964*, 234
Gravity, The Physiological Effects of, ii, 334
Growth of the Human at the Time of Adolescence, i, 308
—, Some Chemical Aspects of Abnormal, iv, 54
Growth Hormone, Human, *1963*, 36
— of the Anterior Pituitary Gland, iii, 137

Haemoglobins, Human, ii, 287
Haemolytic Anaemia, Acquired, vii, 59
Haemophilia and Christmas Disease, vii, 80
Haemopoietic Cells, Transplantation of, *1962*, 327
Haemorrhage: The Effects of Haemorrhage on the Cardiovascular System in Man, i, 79
Health, The Contribution of Science to the Practice of, i, 267
Heart-lung Machines, The Principles of, vi, 85
Heart, Regulation of the Output of the Mammalian, *1965*, 164
Hemispherectomy and the Localization of Function, iii, 93
Hepatic Coma, Biochemical Studies in, viii, 407
—, The Treatment of, v, 380
High Altitude, Man at, *1964*, 32
Histochemical Techniques in Study of Dental Tissues, *1966*, 146
Histochemistry and its Application to the Basic Sciences, iv, 358
—: Enzyme Histochemistry and its Biochemical Alignments, *1962*, 215
History, Science and, iii, 1
Hypercorticalism, The Pathological Physiology of the Human Adrenal Cortex and its Relationship to, *1964*, 15
Hyperoxaluria, viii, 358
Hyperparathyroidism, Parathyroid Function and Primary, *1963*, 266
Hypersensitivity: Does Hypersensitivity cause Disease?, *1964*, 189
—, Laboratory and Clinical Finding in, iv, 123
Hypertension, Pulmonary, viii, 226
—, Some Biochemical Aspects of, vii, 182
Hypertensive Encephalopathy, The Significance of, viii, 256
Hypoglycaemic Agents, Oral, *1962*, 277
Hypothalamic Control of the Cardiovascular Responses in Fear and Rage, *1965*, 217
Hypothermia: Experimental Hypothermia in Animals, v, 19; *see also* Cooling of the Whole Organism

Immune Cellular Reactions, *1961*, 31
Immunization: Active Immunization against Measles, *1964*, 151
Immunity, Antiviral, iii, 259
—, Some Aspects of Antibacterial, i, 192
Immunoglobulins, Disturbances of, *1966*, 106
Immunological Responses in Man to Dusts of Vegetable Origin, *1964*, 203
— Tolerance, Specific, *1965*, 124
Inflammation, Mediators of the Vascular Phenomena of, viii, 198
Injury, The Biochemical Responses to, *1961*, 172
—, Reaction of Enzymes to, iv, 224
—, The Reaction of Epithelium to, *1964*, 276
—, Tissue Response to, i, 130
—: *see also* Trauma
Interfacial Films and Cell Membranes, *1965*, 270
Interference Microscope: The Use of the Interference Microscope in Biological Research, v, 105

Interferon, The Antiviral Action of, *1961*, 21
Intersexuality, viii, 48
Intestinal Smooth Muscle, Physiology and Pharmacology of, vii, 374
Intestine, Absorption from the Small, *1963*, 171
Iron Metabolism, Some Aspects of, viii, 329
Isoenzymes, *1964*, 334
Isotopes: The Relation of Radiation Dose to Skeletal Damage from Bone-seeking Isotopes, *1961*, 47

Kidney Function, Long-term Dialysis as Substitute for, *1966*, 201
—, Studies of Normal and Pathological Physiology of the, i, 143

Larynx: The Comparative Anatomy of the Larynx with particular refernce to the Functions of the Organ in Man, iv, 332
Leukaemia, Contributions of Chemotherapy to the Study of, *1962*, 152
— Research, Epidemiological Methods in, *1961*, 74
Lifespan, Radiation and, *1965*, 91
Limb Blood Flow, Nervous Control of, *1966*, 260
Lipids, Some Factors affecting the Level and Composition of Plasma, *1961*, 262
Liver Cancer, Biochemical and Pathological Mechanisms in Experimental, *1962*, 172
— Failure, *1961*, 216
— Regeneration, *1961*, 236
Low Temperatures, Preservation of Living Cells at, ii, 250
Lung Function Tests, vi, 58
Lungs at Birth, *1965*, 202
Lymphoid Tissue, Some Problems of, vii, 127

Magnesium Metabolism, *1963*, 216
— Metabolism in Man, *1961*, 270
Malaria Parasite, The Life History of the, ii, 323; *see also* Antimalarial Drugs
Mammalian fast and slow skeletal Muscle, *1965*, 186
— Heart, Regulation of the Output of, *1965*, 164
Mammary Function, Recent Light on, i, 39
Measles, Active Immunization against, *1964*, 151
Medical Research, The Scientific Approach to, i, 1; *see also* Clinical Science, Research, Scientific Research
Mental Disorder, Epidemiology and, *1965*, 299
Metabolic Processes, The Steering of, v, 126
Metabolism, Bile Pigment, *1963*, 197
—, Magnesium, *1963*, 216
Micro-organisms, The Nutrition of, v, 285
—, Physiological Adaptation in, *1962*, 34
Molecules, The Shapes of, *1964*, 317
Multiple Sclerosis, The Biochemistry of, *1961*, 283
Muscle, Living, v, 297
—, Mammalian fast and slow skeletal, *1965*, 186
Muscular Contractions, Proteins in, v, 314

Nasal Circulation, Physiology of, v, 451
— Development, Aspects of Facial and, *1963*, 99
Nervous System: Experiments on Single Neurones within the Central Nervous System, *1964*, 81
— and its Disorders, The, iii, 78
—, The Structure and Function of the Supporting Cells of the, *1962*, 294
Neurochemistry, vi, 168
Neurological Mechanisms and Control of Sexual Behaviour, *1965*, 316
Neuromuscular Blocking Substances, Factors Influencing the Action of, vi, 208
Neurones: Experiments on Single Neurones within the Central Nervous System, *1964*, 81
New-born, Anoxia of the, vii, 348
Nose, Ear and Throat, Electron Microscopy in the Pathology of the, *1963*, 302

Nuclear Sexing and Human Chromosome Anomalies, *1962*, 124
Nucleic Acids, *1965*, 17
Nucleoproteins in Cell Structure, IV, 34
Nutritional Assessment of the Individual, I, 364

Oesophagus and Cardia, The Physiology of the Lower, V, 398
Organs, Homotransplantation of, VI, 13
Otological Practice, Some Remarks upon the Scientific Basis of, VII, 1
Output of Mammalian Heart, Regulation of, *1965*, 164
Oxygen, Tissue Requirements, *1966*, 300

Pain, Causes of, VI, 143
Parathyroid Function and Primary Hyperparathyroidism, *1963*, 266
Parturition, The Physiology of, II, 235
Penicillins, Semi-synthetic, *1964*, 169
Phosphatase: Acid and Alkaline Phosphatase in Disease, III, 187
Phosphates in Vital Processes, *1964*, 1
Phospholipids, VI, 217
Physiology, Medicine and the Basic Sciences, *1961*, 1
—: Pathological Physiology of the Human Adrenal Cortex and its Relationship to Hypercorticalism, *1964*, 15
Pia Mater and the Choroid Plexuses, Some Aspects of the Anatomy of the, *1963*, 125
Pituitary Gland and its Blood Supply, The, *1963*, 83
Plasma Aproteinaemias, Primary Protein Deficiencies with special reference to the specific, V, 165
— Lipids, Some Factors affecting the Level and Composition of, *1961*, 262
— Proteins, The, IV, 112
Poliomyelitis, VI, 363
—, VII, 426
—, Vaccination against, VIII, 132
Polysaccharide Chains, Regular Folds in Protein and, *1962*, 203
Polysaccharides, The Synthesis and Degradation of, VI, 233
Porphyrias, The Chemistry of the, IV, 74
Posterior Pituitary Hormones, Release and Activity of, IV, 286
Posture and Electromyography, *1963*, 158
Potassium Deficiency, VII, 451
—: Transport of Sodium aud Potassium across Cell Membranes, *1966*, 217
Pregnancy, Isotopes in the Study of Problems of, V, 90
Protein Malnutrition, IV, 145
— and Polysaccharide chains, Regular Folds in, *1962*, 203
—Ribbons and Sheets, V, 217
Puberty, Age of, and Educational Implications, *1966*, 91
—, The Timing of, *1963*, 53
Pulmonary Blood Flow, Distribution of, *1966*, 279
Psychiatry, Social, VI, 116
Psychopathology, Experimental, III, 372

Quinidine: The Action of Quinidine, Acetylcholine and Anaphylaxis interpreted from Simultaneous Records of Contractions and Intracellular Potentials in the Heart, *1961*, 302

Radiation and Lifespan, *1965*, 91
— Biological Action of, I, 379
— Injury to Bone, IV, 196
—, Recovery from the Lethal Effects of, V, 439
—: The Influence of Oxygen on the Response of Cells and Tissues to Ionizing Radiation, VII, 314
— as a Toxic Agent, VIII, 65
—: The Relation of Radiation Dose to Skeletal Damage from Bone-seeking Isotopes, *1961*, 47
—, Tumour Anoxia and the Response to, *1965*, 74
Radiation-Induced Genetic Changes, *1964*, 234
Radioactive Tracer Methods in Clinical Medicine, *1964*, 250
Radioactivity, Environmental, and its Biological Significance, *1966*, 16
Radiotherapy of Cancer, Fundamental Aspects of the, *1962*, 305
Rage, Hypothalamic Control of the Cardiovascular Responses in Fear and, *1965*, 217

Red Blood Cells, The Life-Span of, ii, 692
—, Viability and Survival of, viii, 269
Renal Failure, Retention of Aromatic Compounds in Acute, *1963*, 235
Reproduction, External Factors in Mammalian, *1963*, 25
Research, The Adventure of, *1961*, 324
Rheumatoid Factor, *1966*, 128

Science and History, iii, 1
Science, Ethical Basis of Medical, *1966*, 1
Scientific Research, Hypothesis and Speculation in, v, 1; *see also* Clinical Science, Medical Research, Research
Sensation, Physiological Problems in the Central Pathway for Somatic and Visceral, vii, 398
—, The Physiological Basis of Visual, i, 116
Sensations, Peripheral Mechanisms Responsible for, vii, 415
Sex-chromosome Anomalies, *1965*, 141
Sex Hormones, The Physiological Actions of the, iii, 169
Sexing: Nuclear Sexing and Human Chromosome Anomalies. *1962*, 124
Sexual Behaviour, Neurological Mechanisms and Control of, *1965*, 316
Sherrington, Cajal and, vi, 100
Silicosis, ii, 108
Skeletal Damage: The Relation of Radiation Dose to Skeletal Damage from Bone-seeking Isotopes, *1961*, 47
Skeletal Muscle, Mammalian and fast slow, *1965*, 186
Skeletal Tissue in Culture, The Physiology of, vi, 28
Sleep: Physiology of Sleep accompanying Dreaming, *1964*, 102
Small Intestine, Absorption from the, *1963*, 171
Sodium and Potassium, Transport of, across Cell Membranes, *1966*, 217
Staphyococcal Infection in Hospitals, Epidemic, viii, 165
Staphylococci, Contrasting Patterns of Drug Resistance in Tubercle Bacilli and, *1962*, 77
— Endemic in Hospitals, *1966*, 157
—: Resistance of Staphylococci to Antibiotics, vi, 333
Stilboestrol, *1965*, 1
Stress and Thyroid Activity, iii, 154
Strokes, The Cerebral Circulation with special reference to, *1964*, 67
Supporting Cells of the Nervous System, The Structure and Function of, *1962*, 294

Temperature Regulation, Disordered *1965*, 239
Throat: Electron Microscopy in the Pathology of the Ear, Nose and Throat, *1963*, 302
Thrombosis, Structural Aspects of, *1964*, 55
Thymus, Functions of the, *1964*, 218
Thyroid and Central Nervous Development, *1966*, 317
Thyroid Hormone, The Biliary Excretion of, vi, 313
Tissue Permeability: Substances that increase Tissue Permeability and their Relation to Infection and Fertilization, iv, 266
— Oxygen Requirements, *1966*, 300
— Repair, iii, 52
— Response to Injury, i, 130
Tissues: Functional Aspects of the Vascularization of Tissues, *1963* , 76
Toxicity, Selective, ii, 14
—, The Elucidation of, v, 244
Toxicology, Industrial, v, 262
Trace Elements in the Brain, *1965*, 37
Tracer Methods: Radioactive Tracer Methods in Clinical Medicine, *1964*, 250
Trachoma and Inclusion Blenorrhoea, *1962*, 359
Transplantation of Haemopoietic Cells, *1962*, 327
Trauma, The Electrolyte and Metabolic Response to, iv, 182; *see also* Burns, Injury
Tubercle Bacilli: Contrasting Patterns of Drug Resistance in Tubercle Bacilli and Staphylococci, *1962*, 77
—, The Epidemiology of, *1963*, 319
Tuberculosis, Experimental Approach to the Problems of Resistance to, vii, 203

Tumour Anoxia and the Response to Radiation, *1965*, 74

Ultrasonics in Clinical Diagnosis, *1966*, 38
Urine, Dilute and Concentrated, *1966*, 192

Vaccines, Virus, *1965*, 282
Vascularization of Tissues, Functional Aspects of the, *1963*, 76
Vegetable Dusts: Immunological Responses in Man to Dusts of Vegetable Origin, *1964*, 203
Virus Adaptability in Relation to Human Disease, II, 46
— Infection, Antibodies and Immunity to, I, 230
— —: Factors in the Transmission of Virus Infections from Animals to Man, *1964*, 125
— Research, The New Look in, IV, 18
— Vaccines, *1965*, 282
Viruses as the Cause of Disease, I, 214
—, The Electron Microscope in the Study of, VI, 379
—, The Growth of, VII, 223
Viscera, The Effects of Distension of, IV, 305
Visual Sensation, The Physiological Basis of, I, 116
Vitamin A, V, 143